

Self-assessment in General Practice Vocational Training

Susan Peta Goldman
B Ed, Grad Dip Soc Sci

Submitted in fulfilment of the requirements for the Degree of
Doctor of Philosophy

School of Human Life Sciences
University of Tasmania
December 2007

Candidate Declaration

I certify that the thesis entitled *Self-assessment in General Practice Vocational Training* submitted for the Degree of Doctor of Philosophy, University of Tasmania contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis. To the best of my knowledge and belief no material has been previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Susan Peta Goldman

19 December 2007

Statement of Access

This thesis may be made available for loan and limited copying in accordance with the *Copyright Act 1968*.

Publications and Presentations

The following published works and presentations arose from the self-assessment study in this dissertation:

Goldman, S. (1998). National medical educator and supervisor focus group data. Unpublished transcript.

Goldman, S., Jasper, A., & Wellard, R. (1998). *A pilot study to evaluate registrar views on outcomes of the RACGP Training Program*. Melbourne: Royal Australian College of General Practitioners.

Goldman, S. (1999a). *A study of registrar views to evaluate outcomes of the RACGP Training Program*. Melbourne: Royal Australian College of General Practitioners.

Goldman, S. P., Swerissen, H., Jasper, A. A., & Wellard, R. F. (1999). Self-reporting of competence by general practitioners with vocational training. *Focus on Health Professional Education: A Multidisciplinary Journal*, 1(2), 50–59.

Goldman, S., Jasper, A., & Wellard, R. (2000). *A study of registrar views to evaluate outcomes of the RACGP Training Program*. Melbourne: Royal Australian College of General Practitioners.

Goldman, S. (2000b). *A pilot study to evaluate supervisor views of general practice vocational training*. Melbourne: Royal Australian College of General Practitioners.

Goldman, S. (2000d). *Supervisor views on aspects of the RACGP Training Program*. Melbourne. Royal Australian College of General Practitioners.

Goldman, S., Cooper, B., Jasper, A., & Wellard, R. (2001). *The effects of gender and locality on registrar views of the Training Program*. Melbourne: Royal Australian College of General Practitioners.

Goldman, S., Jasper, A., & Wellard, R. (2001). *A further study of registrar views to evaluate aspects of general practice vocational training*. Melbourne: Royal Australian College of General Practitioners.

Goldman, S., & Jasper, A. (1998). *GP Training towards the 21st Century*. Paper presented at the RACGP National Medical Education Conference, Canberra.

Goldman, S. (2000a). *A comparison of registrars under new training arrangements with those who were not*. Paper presented at the National Medical Educator Conference, Sydney.

Goldman, S. (2000c). *Does general practice vocational training achieve its objectives?* Paper presented at the Royal Australian College of General Practitioners Annual Convocation, Sydney.

Cooper, B., Goldman, S., Jasper, A., & Wellard, R. (2001b). *Deconstructing feedback*. Paper presented at the National Medical Educator Conference, Townsville.

Acknowledgements

A large team of individuals provided support and encouragement during the completion of this project. Professor Madeleine Ball provided supervision and ongoing support for my research. With her encouragement, the research and the dissertation came to fruition. Associate Professor Dominic Geraghty also deserves a thank you for his help during the last stages of the submission of the thesis. I would also like to acknowledge the efforts of Dr Rod Wellard. In addition to providing support for the research on behalf of the Royal Australian College of General Practitioners, Rod provided his own insights from an educational perspective. Amy Jasper, also from the Royal Australian College of General Practitioners, played an important part in providing ongoing support for the research.

I was fortunate also to have the help and assistance of Dr Chris Peterson. As well as statistical assistance and comments on drafts, Chris provided invaluable support and inspiration. Heather Gardner was another valuable support person who patiently proof read the dissertation.

There were 1500 trainees over three years, who contributed their valuable time in completing questionnaires and providing their views about general practice vocational training. Without the trainees' preparedness to declare frankly their own strengths and weaknesses, there would not have been a self-assessment study. A thank you is also due also to Glynis Ramsay whose administrative assistance was outstanding. A dedicated crew of telephone callers, largely responsible for the high response rates for the three surveys, also deserve my thanks.

Finally, I thank Dr Stephen Goddard who has supported and encouraged me, not only on this journey, but also in life.

Contents

Candidate Declaration	iii
Statement of Access.....	iv
Publications and Presentations	v
Acknowledgements	vii
Abstract	xvi
Chapter 1: Introduction.....	1
The Beginning of the Evaluation of General Practice Training	1
Context.....	2
Evaluation Approach of the Outcomes Evaluation Unit	3
Origin of the Self-assessment Study.....	4
Research Study Strategy and Questions	5
Significance of the Self-assessment Study	7
Thesis Overview	9
Chapter 2: General Practice Vocational Training.....	11
Origins	11
Influences on the Curriculum	17
The Development of Australian General Practice Competencies	18
Summary of the Origins of General Practice Vocational Training	22

Chapter 3: Perspectives of Learners.....23

The Field of Self-assessment Studies and General Practice	
Vocational Training.....	23
Self-assessment and Educational Evaluation	26
Attitudes to Training	26
Summary of Self-assessment and Educational Evaluation	35
Validation Studies.....	37
Questions Around Validation Studies and Self-assessment	38
Interpretations Around Difficulties of Validation Studies....	41
Summary of Validation Studies.....	43
Interventions to Teach Self-assessment.....	44
Summary of Interventions to Teach Self-assessment.....	45
Current Frameworks for Self-assessment	45
The Need for Self-assessment in Medical Education	46
Benefits of Self-assessment	47
Assessment as a Collaborative Process	48
Measurement Rigour versus Educational Relevance.....	49
Changing Methodologies.....	50
Summary of Current Frameworks for Self-assessment.....	51
Towards a Model for Evaluating Training	51

Summary of the Field of Medical Vocational Training and Self- assessment Studies.....	52
Chapter 4: Perspectives of Medical Educators and Supervisors.	54
Origins of Experiential Learning and Self-assessment.....	55
Uncertainty in Medicine.....	58
The Critique of Traditional Models of Learning and Teaching in Medical Education	60
Influential Theories in Australian General Practice Training	62
Reflective Practice.....	64
Theoretical Basis for Reflective Practice	65
Reflection-in-action	66
Summary of Reflective Practice	71
Adult Learning	71
The Need to Know	72
Learner's Self-concept.....	74
The Role of the Learner's Experience—Emphasis Upon Experiential Techniques and Peer Learning	77
Readiness to Learn.....	78
Orientation to Learning.....	79
Motivation.....	80
Conclusion on Influential Theories on General Practice Vocational Training.....	81

Chapter 5: A Study of Aspects of General Practice Vocational Training.	84
Aims	84
Design Considerations.....	85
Research Strategy	88
Hypotheses.....	91
Ethics	93
Method.....	93
Procedure	98
Questionnaire Development.....	100
Sample.....	103
Recruitment	103
Questionnaire Data Analyses	104
Comparison of Questionnaire and Examination Data.....	107
Summary of the Design of the Self-assessment Study	107
Chapter 6: Study Findings.....	109
Characteristics of Respondents.....	109
Self-assessed Competence	109
Domain One: Communication Skills and the Patient-Doctor Relationship	112
Longitudinal Results on Domain One	115

Domain Two: Applied Professional Knowledge and Skills	116
Longitudinal Results on Domain Two	121
Domain Three: Population Health and the Context of General Practice.....	122
Longitudinal Results on Domain Three	125
Domain Four: Professional and Ethical Role	126
Longitudinal Results on Domain Four.....	129
Domain Five: Organizational and Legal Dimensions	129
Care of Particular Patient Groups.....	133
Longitudinal Results on the Care of Particular Patient Groups	139
Care of Patient Presentations.....	141
Longitudinal Results on Patient Presentations.....	145
Satisfaction	145
Longitudinal Findings on Satisfaction	149
Satisfaction and Self-assessed Competence	149
The Relationship Between Self-assessments and Examination Results	154
Self-assessed Competence.....	154
Satisfaction	156
Findings in Relation to Study Hypotheses	156

Chapter 7: Discussion of the Self-assessment Study.....	158
Reflections	158
Study Research Questions and Contribution.....	159
Methodology	161
Limitations	163
Perspectives of Learners and Self-assessed Competence	
Findings	165
Stage of Training	167
Gender	169
Locality	170
Workload	171
Longitudinal Results	173
Perspectives of Learners and Satisfaction Findings	173
Perspectives of Educators and Study Findings.....	178
Conclusion	182
References	185
Appendix A.....	200
Appendic B.....	232
Appendix C.....	269
Appendix D.....	271

List of Tables

Table 1: Characteristics of current and former trainees.....	110
Table 2: Further characteristics of current and former trainees.....	111
Table 3: Stage of training by Domain one.....	114
Table 4: Workload by Domain one	115
Table 5: Longitudinal findings for Domain one	116
Table 6: Stage of training by Domain two.....	118
Table 7: Stage of training by referrals	119
Table 8: Workload by Domain two.....	119
Table 9: Workload by further items in Domain two	120
Table 10: Longitudinal findings for Domain two	121
Table 11: Stage of training by Domain three	123
Table 12: Stage of training by public health items	124
Table 13: Longitudinal findings for Domain three	125
Table 14: Stage of training by Domain four	127
Table 15: Workload by Domain four	128
Table 16: Longitudinal findings for Domain four	129
Table 17: Stage of training by Domain five	131
Table 18: Workload by Domain five	132

Table 19: Stage of training by particular patient groups	134
Table 20: Stage of training by women to indigenous people	135
Table 21: Stage of training by people with mental illness to chronic conditions.....	136
Table 22: Workload by particular patient groups.....	138
Table 23: Workload by women.....	139
Table 24: Workload by men to people with chronic conditions	140
Table 25: Workload by people with disabilities.....	141
Table 26: Longitudinal findings for particular patient groups.....	142
Table 27: Predictors of self-assessed competence for year one.....	151
Table 28: Predictors of self-assessed competence for year two	153
Table 29: Predictors of self-assessed competence for year three	154

Abstract

- Objectives:*** This dissertation explores the utility of self-assessment for trainee general practitioners in Australia and examines the implications of self-assessment for skills acquisition by reflecting upon educational theory and research.
- Aims:*** The study at the core of this dissertation aims to evaluate general practice vocational training from the perspective of learners using self-assessment. Using a model developed from the self-assessment literature, the study examines the extent to which trainees view themselves as competent in key curriculum areas. Research questions consider whether participation in general practice training results in a measurable increase in self-assessed competence in the key areas of general practice training, and to what extent trainees are satisfied with the program. Also considered is the relationship between the two concepts of satisfaction and self-assessed competence, and the correlation between examination results and self-assessments.
- Method:*** Qualitative analysis of empirical research on self-assessment studies in medical education and selected theories of adult and experiential learning produced a model with which to evaluate aspects of general practice vocational training from the perspective of the learner. A national sample of trainees of the Royal Australian College of General Practitioners participated in a three-year longitudinal panel study from 1999 to 2001.

Results:

Six hundred and twenty-two trainees completed the survey in year one, 549 trainees in year two and 472 trainees in year three. The sample mirrored that of the population of general practice trainees. A sample of 161 trainees participated in all three surveys and their results on competence and satisfaction were compared over time. The model tested provided information about the significance and role of self-assessment to evaluate aspects of general practice vocational training from the perspective of the learner. It established that participation in general practice vocational training resulted in an increase of self-assessed competence in areas of the curriculum, and that trainees were generally satisfied with their training experiences. The variables of stage of training, gender, locality and workload provided insights into the utility of trainee self-assessments, particularly in relation to their strengths and weaknesses. The results suggest that as they progress, the longitudinal cohort of trainees viewed themselves as improving an average of nine percent in their self-assessed competence rating in the five domains of general practice, seven percent for the care of particular patient groups and 11 percent for the management of particular patient presentations. This pattern of consistent differences, in the same individuals over time, supported the idea that a process is at work whereby trainees' confidence improved as they progress through training.

Implications: The baseline data emerging from this study provide findings that might usefully contribute to the debates surrounding future training needs and self-assessment, and the enhancement of the conditions around the delivery of general practice vocational training. This is particularly pertinent to the environment of general practice vocational training, in which there are multiple training providers, and to the development of continuing medical education.

Chapter 1: Introduction

This dissertation examines the role of self-assessment in general practice vocational training and its implications for fostering lifelong learning. The Introduction presents a short history of the establishment of evaluation at the Royal Australian College of General Practitioners. Some contextual information around the process of evaluation of general practice vocational training provides the origins of the self-assessment study at the core of the dissertation. The Introduction also presents the research strategy and questions, their significance and a description of how the dissertation is structured.

The Beginning of the Evaluation of General Practice Training

Training for General Practitioners in Australia began in 1973, with the establishment of the Family Medicine Programme by the Royal Australian College of General Practitioners. Before this time, general practice was not a highly regarded discipline. Many trainees entered the profession without a structured training program. Although by 1982, the Family Medicine Programme included a certificate of satisfactory completion, it was not until the advent of vocational registration in 1991 that the distinction was made between the monitoring of practice behaviour and evaluation of the program. With the advent of vocational registration and summative assessment through the Fellowship Examination, the program (by then the Royal Australian College of General Practitioners Training Program) became subject to the evaluation of educational effectiveness with the setting up of an Evaluation Directorate, headed by Dr John Docker. Later in 1995, funding of the Training Program, in the form of the Commonwealth Funding Agreement, became contingent on the implementation of merit-based selection, recognition of prior learning, a formal curriculum, and

a substantial program of evaluation. In response, the Royal Australian College of General Practitioners embarked on both process evaluation, through the Quality Improvement Unit, and outcomes evaluation, through the Outcomes Evaluation Unit. The author began work with the Royal Australian College of General Practitioners in 1997, and as a member of the Outcomes Evaluation Unit conducted the self-assessment study central to this dissertation between 1999 and 2002.

Context

Several key aspects of the work of the Royal Australian College of General Practitioners Outcomes Evaluation Unit provided the backdrop to this dissertation. Ideally, program evaluation should be a part of program development (Stufflebeam, Foley, Gephart et al., 1971); but it is by no means unusual for the two processes of program development and evaluation to occur asynchronously. In addition, the establishment of evaluation is not necessarily a smooth or easy process. In this particular instance, there had been two attempts to introduce an evaluation program for general practice vocational training before the establishment of the Outcomes Evaluation Unit. Most difficulties in the attempts to introduce evaluation stemmed from the different management structures and philosophies of the Royal Australian College of General Practitioners and the Training Program. These difficulties might have been due different understandings about the tasks of general practitioner certification, and the enhancement of the general practice vocational training experience. To avoid the prospect of past difficulties recurring in the third effort to introduce the evaluation of training, the Royal Australian College of General Practitioners structured the Outcomes Evaluation Unit to work both independently from it, and collaboratively with it, through a representative advisory group. In addition, the funding source (as part of the Training Program Grant from the Commonwealth Department of Human Services and Health) ensured that evaluation was a

mandated activity, rather than a discretionary one. This placed the Outcomes Evaluation Unit in a fortunate position whereby much of the effort could be devoted to the actual practice of the evaluation. These arrangements provided a stable environment for the research and evaluation activities over the period of existence of the Outcomes Evaluation Unit from 1996–2002.

Evaluation Approach of the Outcomes Evaluation Unit

As its first task, the Outcomes Evaluation Unit developed the Five-year Strategic Plan for Evaluation of Outcomes of the Royal Australian College of General Practitioners Training Program 1998–2002. The basis of the plan was a conceptual framework of efficiency, effectiveness, and appropriateness (as set down in the funding agreement with the Commonwealth Department of Health and Family Services). This framework and strategy has been frequently applied to the area of performance audit (Victorian Auditor-General's Office, 2003). Essentially, the appeal of such an approach lies in its simplicity and the way in which it accommodates the needs of multiple stakeholders. Some of these stakeholders included government officials, and the senior management of the Training Program and the Examination Department of the Royal Australian College of General Practitioners, as well as academic medicine specialists, medical educators, supervisors and consumer advocates. All of these stakeholders were represented on the advisory committee that had oversight of the Outcomes Evaluation Unit (Royal Australian College of General Practitioners, 1998).

The conceptual basis underlying the plan was that the triangulation of findings might provide insights into the benefits and shortfalls of the program. During its existence, the Outcomes Evaluation Unit conducted a large range of projects. The studies included analyses of trainee satisfaction, self-assessed competence, educational activities, consultation and referral patterns and behaviour compared to other practitioners (Britt, Miller, Sayer, & Sedgwick, 1999; Britt, Miller, &

Valenti, 2001; Cooper, Goldman, Jasper, & Wellard, 2001a; Cooper, Jasper, Goldman, & Wellard, 2001c; Goldman, 1999; Goldman, Jasper, & Wellard, 2001b; Goldman, Swerissen, Jasper, & Wellard, 1999; Prideaux, Henry-Edwards, Middleton, & Lyons-Reid, 1999). Some preliminary work was also conducted on patients' views of their general practitioners (Brownlea & Greco, 1998); but this proved to be both a methodologically difficult and costly area in which to pursue evaluation. Had the Outcomes Evaluation Unit continued beyond the period of the Five-year Strategic Plan, it would have been possible to implement a range of program enhancements to training. At that stage, it would have been appropriate to reconsider the work on patients' views of trainee general practitioners.

Origin of the Self-assessment Study

There is some precedent in the medical education literature for the use of multiple perspectives in educational evaluation, with the work of Tamblyn (1994). She posed a useful framework for understanding medical education and credentialing to protect the public from substandard practitioners, by describing the relationship between quality practice, performance, and patient outcomes in Canada. This framework provided a starting point with which to establish research questions for the self-assessment study at the core of this dissertation.

The problem at issue was an examination of the extent to which trainees viewed general practice vocational education as meeting their needs, and how self-assessment might foster reflective practice. To accomplish this, the self-assessment study at the core of this dissertation concentrated on a model developed from the findings of Arnold, Willoughby and Calkins (1985). The expansion of this model provided an opportunity to gather data that might be used to enhance the educative process.

This self-assessment study is not about general practitioner competence, competence measurement, or ways to address problems pertaining to a lack of competence. Nor does the study suggest that self-assessment would ever be a substitute for the certification of a general practitioner as competent. This task is appropriately the province of those who certify general practitioners as competent through summative examinations and the monitoring of practice behaviour and continuing medical education. In particular, the self-assessment study does not deal with elements of practice such as consumer health outcomes, for studies in this area are both methodologically and logistically difficult to conduct, and outside the scope of a doctoral project in the context described here.

Research Study Strategy and Questions

One of the enduring values articulated in Australian general practice vocational training is that of lifelong learning. It is valued across all adult-learning contexts. Acknowledged as a critical aspect of mindful practice and the ability to be prepared for the next situation, lifelong learning is particularly important in medicine (Epstein, 1999). This is partly because of the degree to which autonomy and uncertainty are associated with the discipline of medicine. As such, the importance of lifelong learning is predicated on the recognition of the need continually to identify new skills and strategies. The pathway to this awareness (more recently referred to as mindful practice) is self-assessment. This dissertation explores the utility of self-assessment for general practice trainees and examines the implications of self-assessment for skills acquisition by reflecting upon educational theory and research. The focus of the thesis is both an appraisal of the use of self-assessment for those engaged in general practice vocational training, and an analysis of a longitudinal study conducted with trainees of the Royal Australian College of General Practitioners, to examine the perspective of the learner. Expanding upon a model developed from previous findings in the literature, the self-assessment study measures the extent to which

trainees assess themselves to be competent in five domains of general practice in the curriculum, and report satisfaction with their training as they progress. The study also considers the relationship between lifelong learning and self-assessment. Specifically, the study evaluates aspects of general practice training from the perspective of the trainee by answering these research questions:

- Does participation in general practice vocational training result in a measurable increase in self-assessed competence in the five domains of general practice in the curriculum?
- What differences are there between male and female practitioners in their self-assessed competence ratings?
- What differences are there between urban and rural practitioners in relation to their self-assessed competence ratings?
- What is the appropriate number of patients for trainees to manage as indicated by self-assessed competence ratings?
- To what extent do self-assessed competence ratings of a matched cohort of trainees increase as they progress?
- To what extent are general practice trainees satisfied with their training experiences within the program?
- Does satisfaction increase as trainees progress through training?
- What differences are there between males and females in their satisfaction ratings?
- What differences are there between urban and rural practitioners in their satisfaction ratings?

- What is an appropriate number of patients for trainees to manage as indicated by their satisfaction ratings?
- To what extent do satisfaction ratings of a matched cohort of trainees increase as they progress?
- What is the relationship between satisfaction and self-assessed competence ratings?
- To what extent do self-assessed competence ratings and examination results correlate with one another?
- What might be the role and significance of self-assessment in relation to general practice vocational training?

These questions are addressed in Chapter 7 in a discussion of the important findings of the study and their implications for education and self-assessment practices.

Significance of the Self-assessment Study

The rise of general practice as a discipline and a profession has been met with the realization that undergraduate medical education cannot fully prepare graduates to engage in independent, general practice (Commonwealth Department of Health and Family Services, 1996). Arising from their autonomous status, is the need for practitioners to exercise self-direction not only in their practice, but also in the pursuit of lifelong learning (Royal Australian College of General Practitioners, 2002). In this context, a prerequisite for lifelong learning is the recognition of the areas in which general practitioners need to develop skills; in other words, the capacity to engage in self-assessment as a form of reflective practice.

The self-assessment study at the core of this dissertation provides insights into the value of general practice training from the perspectives of learners. First, the field of self-assessment for educational evaluation is examined using a typology derived from the literature that distinguishes between the different uses of self-assessment. Second, through the linkage to the use of self-assessment in relation to lifelong learning, the self-assessment study makes further contributions by examining the utility of self-assessment. Third, the findings of the self-assessment study potentially foster efforts to enhance the delivery of training by emphasizing educational processes that arise from the literature on the perspective of the educator. Fourth, the exploration of self-assessment contributes to new knowledge by providing baseline data on self-assessed competence and satisfaction. The longitudinal aspects of the study and particularly the ability to follow a small cohort of trainees for three years provide insights into possible influences on trainee self-assessed competence and their implications for the purpose of future program development and enhancement. In summary, this dissertation examines the role of self-assessment in an everyday general practice environment, and its implications for skills acquisition and lifelong learning.

The system of training that existed at the outset of the project has changed from being delivered by the Royal Australian College of General Practitioners to one under which training is delivered by many different consortia under the management of General Practice Education and Training, a government-owned company, acting as the fund holder. The recommendations developed will be of relevance to future evaluation of the multiple general practice vocational training providers and to the Royal Australian College of General Practitioners in its provision of Quality Assurance and Continuing Professional Development.

Thesis Overview

Chapter one describes the purpose of the dissertation and the significance of the study conducted with trainees of the Royal Australian College of General Practitioners. It presents a brief history of the evaluation of general practice vocational training providing the context, origins and central research questions of the study at the centre of the dissertation.

Chapter two provides a history of the rise of the discipline of general practice in relation to vocational training and lifelong learning, key events surrounding the program and the curriculum, and arrangements for the evaluation of general practice vocational training.

Chapter three focuses on the literature from previous empirical medical education studies from the perspectives of learners. The examination of the field of postgraduate medical education provides the basis for a typology on the use of self-assessment. These uses include educational evaluation, validity studies on knowledge and skills, interventions around the teaching of self-assessment as well as current frameworks for the implementation of self-assessment in the processes of teaching and learning. The purpose of using this typology is to relate self-assessment to current frameworks that encourage learners to examine their strengths and weaknesses in order to gain feedback. The use of similar collaborative models to encourage learners to solicit feedback and participate in remediation serves to place the learner at the centre of the educational experience.

Chapter four provides a selective exploration of the origins and impact of experiential learning and adult learning theories with particular reference to the contributions of reflective practice from Donald Schön (1987) and adult learning from Malcolm Knowles (1984). At the centre of both of these theories of experiential learning is the emphasis on learner participation. The chapter

explores learner participation in relation to general practice vocational training to develop a context for the self-assessment study.

Chapter five presents the methodology of the self-assessment study. In order to evaluate aspects of general practice training, a model that represents an extension of the literature is used to examine the attitudes of learners. This chapter also explores some of the key questions in the choice of research design, and the major hypotheses on stage of training, gender, locality, workload and self-assessed competence and satisfaction.

Chapter six presents the findings of the self-assessment study, in relation to the areas of the curriculum (such as the five domains of general practice), and the stage of training, gender, locality and the role of self-assessment.

Chapter seven discusses the results of the self-assessment study in relation to the literature. There is also an evaluation of the initial hypotheses and their significance to the role of self-assessment in the process of moving from novice to professional. The conclusion to the chapter provides recommendations in relation to the practical application of self-assessment to learner participation and lifelong learning.

Chapter 2: General Practice Vocational Training

This chapter provides a brief account of the history of general practice vocational training in Australia. It describes some of the key milestones and issues influencing training, key iterations of the program and the different pedagogical views held within the Royal Australian College of General Practitioners. This account, largely reliant on written histories of the program, examines the mode of training adopted, and the development of notions of competency as influences upon the general practice vocational training curriculum. Within this description are insights into the transition from a relatively unstructured and self-directed learning experience to a highly structured, monitored and fully credentialed program.

Origins

Before the 1950s, there was no formal system for specialist training in the discipline of general practice (Bollen & Saltman, 2000). With the rise of specialisation of medicine in the 1950s arose the recognition of the need for a professional guild for general practitioners. In response, the profession established the Australian College of General Practitioners in 1958. By 1962, the Medical Education Committee of the Australian College of General Practitioners Council had formulated an education policy (Commonwealth Department of Health and Family Services, 1998b). By 1973, the organization had changed its name to the Royal Australian College of General Practitioners. It founded the Family Medicine Programme that continued until 1993, when the Royal Australian College of General Practitioners introduced the Training Program.

Virtually from the founding of a general practice vocational training program, the relationship between government through the Commonwealth Department of Health and the Royal Australian College of General Practitioners ensured that the Training Program was contested. The Department of Health began over time to view general practice vocational training as a tool with which to lever changes to the primary care system more generally, as demonstrated by the number of times the Department of Health initiated reviews. Between 1974 and 2001, twelve reviews of the program were conducted (Royal Australian College of General Practitioners, 2002). The recurring interests of various Commonwealth Departments of Health were about general practitioner workforce numbers and distribution, and the ways in which these factors shaped general practice vocational training policy.

Periodically, the interests of the Royal Australian College of General Practitioners diverged from those of the Commonwealth Department of Health. Wilde (1998) recorded an oral history of general practice vocational training, and her interviews with stakeholders of the Royal Australian College of General Practitioners established some pressure points of the period. These revolved around Medibank (Australia's universal health insurance scheme), shortages in the general practice workforce and inter-professional politics around the representation of rural doctors. Wilde provided a linkage between these issues.

In the discussions of 1972 about the proposal for the Royal Australian College of General Practitioners to embark upon the delivery of vocational training, College Council was concerned that its members would perceive the organization as aligned with the Commonwealth government and unable to speak out on policy. After the offer to develop and conduct the Family Medicine Programme was accepted in 1973, the other issue that arose was opposition to summative examinations. This was reflected in the anomalous situation in

which the importance of self-assessment was articulated by medical educators and the voluntary examination was offered to trainees—albeit reluctantly (Royal Australian College of General Practitioners, 1974). From a present day perspective, it might be difficult to understand how this situation prevailed; but according to Wilde's analysis, the government of the day did not oppose the position on certification of practitioners because of the influence of several industrial debates during the period.

At the same time, relatively poor relations prevailed between the College and the Australian Medical Association—the key medical industrial body. This discord between the two organizations assisted the progress of the College in its negotiations with government, which at that stage was urgently required to address the lack of entrants to general practice. Initially, the Australian Medical Association opposed the universal health scheme introduced in the *Health Insurance Commission Act 1973*. At the same time, the Commonwealth government did not want the increased power of the medical industrial body and the Royal Australian College of General Practitioners that the external credentialing of general practitioners would inevitably bring (Wilde, 1998).

By 1982, however, the situation had changed. The Liberal conservative government of the time was keen to hold the Royal Australian College of General Practitioners more accountable for training. The mechanism for this was through the application of a credential upon completion of the Family Medicine Programme. To have no limits on time spent in training had always been problematic for governments interested in increasing workforce numbers. Consequently, the review from Hurley and Cummins (1982) produced the second iteration of general practice vocational training. The new version of the Family Medicine Programme sought to address the open-ended nature of the length of training for the program. After the review, the Department of Health and the

Royal Australian College of General Practitioners reached a compromise position in negotiations whereby the program was to occur over a shorter timeframe. Consequently, Mark II of the Family Medicine Programme, comprised two years of hospital experience, two 13-week terms in general practice, and further general practice experience for those who required it. Learning contracts were to provide the pathway to the endpoint of the program, and a Certificate of Satisfactory Completion of Training became available to provide exemption from some parts of the Fellowship Examination (Fleming, 1993).

There was opposition within the Royal Australian College of General Practitioners to the new plans. Trainees, medical educators, and program management voiced their critique of the proposal on the basis that a summative examination would negatively affect the Family Medicine Programme. This was said to be because of the strong reliance upon self-directed learning of the Family Medicine Programme. At the same time, however, the Family Medicine Programme was not without its internal critics. When Dr Richard Notowny became the State Director in Western Australia after the adoption of Mark II of the Family Medicine Programme, he found that the claims made of the Family Medicine Programme did not always translate to positive educational experiences for trainees (Wilde, 1998). In response, there were efforts to enhance the educational experience for learners with the inclusion of observation of their consultations by external clinical teachers. Changes were afoot, and heralded an era of constraint on the unstructured style of educational innovation so associated with the Family Medicine Programme.

The next, and by far the most significant milestone of the Training Program, heralded the era of vocational registration. Abrahamson (1987) recommended the implementation of summative assessment, and the shortening of training to three years. This made the Fellowship the basic avenue to vocational registration.

General practice reforms of the period provided the context for negotiations involving the Australian Medical Association and the Royal Australian College of general practitioners at the heart of vocational registration. The reforms addressed the quality of the medical workforce, and practice efficiency (Commonwealth Department of Health and Aged Care, 2000). Vocational registration was the outcome of the implementation of standards for trainee general practitioners and contributed to the acceptance of the discipline of general practice.

Some time later, in 1996, the issue of workforce took precedence in negotiations between the Australian Medical Association, the Royal Australian College of General Practitioners and the newly re-titled Department of Health and Family Services. Data presented from the Australian Medical Workforce Advisory Committee suggested a *potential* oversupply of general practitioners (Australian Medical Workforce Advisory Committee, 1996). A restriction upon the number of general practitioners was one way to address concerns at the increasing costs associated with the Pharmaceutical Benefits Scheme. Based on this perceived oversupply of general practitioners, the Commonwealth government implemented a quota on Medicare provider numbers in the same year. The effect of vocational registration was that entry into general practice training became more desirable and competitive. The reaction to these developments among the pool of general practice aspirants was swift. Not only did general practice become recognized as a speciality involving high university scores, but also trainees incorrectly perceived that the restriction on Medicare Provider numbers was within the control of the Royal Australian College of General Practitioners. Trainees' anxieties increased, particularly with the implementation of a mandatory one-year term in rural areas.

An unintended consequence of the annual quota of 400 new recruits was a negative effect on the future supply of rural and outer metropolitan general

practitioners (Hays, Worley, & Coote, 2004). As a response to concerns about the supply of doctors in rural areas, the government implemented another review, *General Practice Education: The Way Forward* (Commonwealth Department of Health and Family Services, 1998a). Sentiments about the program in the review document were unambiguously critical in tone. At the time, the author was in the early stages of her role as a researcher in the Outcomes Evaluation Unit, and can recall that few people within the organization recognized the implications of the criticism for the later developments of general practice vocational training. Although the review found that the Training Program was high quality, transparent and used reliable assessment processes, key doubts remained about workforce numbers and their distribution across metropolitan and rural areas.

The review highlighted other issues around inter-professional politics (Commonwealth Department of Health and Family Services, 1998a). A pressure point seemed to be developing. In 1992, the Rural Doctors Association of Australia and the Royal Australian College of General Practitioners had established the Faculty of Rural Medicine. Three years later, the Royal Australian College of General Practitioners introduced a rural training stream that included one year of training in a rural location. Relations, however, were not smooth within the Royal Australian College of General Practitioners Council and its National Rural Faculty. This split resulted in the formation of the Australian College of Rural and Remote Medicine (Hutton-Czapski, 2000).

After a series of negotiations, the review made recommendations to align training to regions, address the lack of rural posts, and separate accountability for standards and the provision of training. This led to the implementation of a new system in 2002, under which local consortia bid for training contracts and delivered training with standards set by the Royal Australian College of General Practice. One central body, General Practice Education and Training Program Limited, had oversight of the whole system.

In this brief description and history of general practice vocational training, the key events described provide some insights into the climate surrounding general practice vocational training. These aspects and the importance of workforce numbers and their distribution combined to influence policy, frequent program reviews, and key historical milestones. In terms of educational processes, an opposition to summative assessment reflected a preference towards self-directed learning. Each subsequent version of the program became less open-ended, and more tightly supervised. Surprisingly, the quality of the program was never the key issue. There was a body of evidence indicating the effectiveness of the Training Program, and it was recognized as comparable with other programs in the United Kingdom, the United States and New Zealand (Hays & Piterman, 2000). The trend in the United Kingdom and the United States was for local delivery of programs overseen by a central board of studies that sets standards. Although Australia was slower to adopt this trend (Weller & Dunbar, 2004), after several reviews, the system of localized delivery was adopted. The self-assessment study at the core of this dissertation examines the program from the perspective of trainees just before this development, and explores some of the theories influencing medical educators.

Influences on the Curriculum

The program of general practice vocational training in Australia was, from an evaluation perspective, a comprehensive one to evaluate. One approach was to regard the program as a mosaic and accept that it was not viable to consider all aspects. For this reason, the curriculum was accepted as a fundamental part of the program and treated as an educational intervention. Although the field of methods for the development and measurement of competencies might involve much investigation, the focus of the self-assessment study at the core of this dissertation is to evaluate general practice training from the learner's perspective. The following section describes the development of Australian general practice

competencies in relation to general practice and nursing research. The focus is not on the actual competencies, but on the methodology for their development in relation to the curriculum. For this reason, the following section is consistent with the previous discussion on the development of the self-assessment study.

The Development of Australian General Practice Competencies

The Australian experience of competence development was not only concerned with clinical competence; but also encompassed preventive care, communication skills in the consultation, and the extent to which the consultation demonstrates collaborative relationships. As reflected in the literature on competence development, the repertoire changed over time to encompass broader skills such as knowledge recall, interpretive skills, problem-solving skills, affective behaviours, perceptual and manual skills, and the attributes of a doctor.

The work of Benner (1984) provides a useful framework for interpreting the development of competencies and the curriculum for general practice vocational training in Australia. In demystifying the process of moving from novice to expert in nursing, Benner analyzed the component parts of nursing care, and identified its domains of competence. Her analysis used an established qualitative analysis methodology to interpret fundamental meanings and content. The idea of founders of general practice vocational training in Australia engaging in interpretative qualitative analysis runs counter to the prevailing ethos of the Training Program; however, both Benner and the program founders sought to capture the day-to-day experience of practising clinicians by capturing the essence of professional practice. Benner made qualitative judgments about common meaning apparent through nurse's narratives using the analysis of common themes. In the case of Australian general practice training, the considerable observational research of Bridges-Webb was the source for a portrait of everyday general practice. One influential study of curriculum

design featured general practitioners recording a week's consultations over several years (Andersen, Bridges-Webb, & Chancellor, 1986). From research based on the International Classification of Morbidity, Bridges-Webb (1976) represented the prevalence of the key patient problems presenting to general practitioners. From this research, the general practice vocational training program curriculum was broadly based on five domains of general practice and a series of priorities with regard to particular patient groups and presenting conditions. The domains of general practice include communication skills and the patient-doctor relationship, applied knowledge and skills, population health and the context of general practice, professional and ethical role, and organizational and legal dimensions (Royal Australian College of General Practitioners, 1999a, 1999b). The Fellowship Examination of the Royal Australian College of General Practitioners was also based upon the research from Bridges-Webb, incorporating a matrix of the five domains of general practice, and particular patient problems and presentations to be sampled in the examination based on their frequency. The connection between actual practice and the curriculum might account for the high content validity of the examination established in subsequent research (Spike, 1999).

The striking feature both of the competencies identified by Benner and those of the general practice vocational training program was the emphasis upon the relationship with the patient, and the teaching, learning and coaching role inherent in this relationship. Equally important to the competencies developed by Benner for nursing, was the priority accorded to particular areas of the general practice vocational training program curriculum. The validity of this priority was evident in an extensive literature review that found that the emphasis on patient-general practitioner communication was warranted in terms of current conceptions of the importance of the active role of patients in their care (Prideaux et al., 1999).

Historical influences on competence development have shaped the progress and sequencing of evaluation of training and the curriculum. The articulated objective of Australian general practice vocational training was that, after training, trainees should become competent and independent practitioners (Royal Australian College of General Practitioners, 2000). Although this has been a long-standing objective of general practice vocational training, the development of defined competencies representative of family practice in the United Kingdom and general practice in Australia, was a relatively recent development. This delay in developing competencies contributed to a delay in efforts to evaluate these programs; hence, the processes of program design and of evaluation are somewhat separate in Australia.

The curriculum for general practice vocational training in Australia was a major component of the educational program. In seeking to evaluate the effectiveness of the program as an intervention, a key question that arose is the extent to which general practitioners achieve competence at the end of training. It is no exaggeration to suggest that the field of competence development and assessment is substantial; but the interest in this dissertation is in the relationship between competence and self-assessment. From a selective analysis of the literature on competence development, several findings relate to the role of self-assessment.

First, competence is an ever-expanding and evolving field. In the 1980s and 1990s, Campbell and Murray (1996), Donnelly, Boltri, Weeks, and Mengel (1994), Fraser, McKinley and Mulholland (1994) are just a few examples of those who no longer depicted competence as a static concept. Instead, they included the broader dimensions of problem solving, consulting and reporting skills, and the demonstration of skills, attitudes and knowledge. Conceptions of competence broadened to embrace non-cognitive skills such as communication skills, for example. With the introduction of self-assessment of competence,

researchers began to use the concept of readiness to practice. This concept was also used to develop recommendations with which to reform curricula (Arnold, Willoughby, & Calkins, 1984; Arnold et al., 1985; Biro, Siegel, Parker, & Gillman, 1993; Blumenthal, Gokhale, Campbell, & Weissman, 2001; Chambers, Wall, & Campbell, 1996). A move away from this concept concentrated on more sophisticated dimensions in the measurement of competence. A substantial review from Epstein and Hundert (2002), for example, proposed cognitive, integrative, context, relationship, and moral dimensions as well as habits of mind as contributing to a holistic picture of competence for medical practitioners. There is scope for researchers to make use of these models more frequently.

Second, the purpose of competence assessment has determined the way in which competence was conceptualised and measured (Allen & Rashid, 1998; Bashook & Parboosingh, 1998; Carlisle, 2000; Fielding, Page, & Schulzer, 1982; Hojat, Veloski, & Borenstein, 1986; Maudsley & Strivens, 2000a; Miller, 1990; Newble, Dawson, Dauphinee et al., 1994; Rethans, van Leeuwen, Drop, van der Vleuten, & Sturmans, 1990; Southgate, 1994; Southgate, Campbell, Cox et al., 2001; Stewart, O'Halloran, Barton et al., 2000; Thomson, 1990). These researchers concluded that summative assessment for the purpose of certification of competence has more of an emphasis on clinical and procedural skills than formative assessment for the assessment of other skill domains. These domains, featured in the self-assessment studies in Chapter three, provide further background to the self-assessment study at the core of this dissertation.

Third, and of most importance to the self-assessment study, is the relationship between objective and subjective assessments of competence. Miller (1990) presents competence as a hierarchical pyramid of skills and abilities, valuing action at the apex, and knowledge at the base. Examinations to drive the system of learning by stipulating clear outcomes for trainees were based on performance-

based assessment. The Fellowship Examination of the Royal Australian College of General Practitioners drew on this pyramid by emphasizing performance-based assessment. This is also the system adopted in the United Kingdom, and the Netherlands. Evidence from Spike and Veitch (1990) and Spike (1999) on the Fellowship Examination in Australia demonstrated the usefulness of this approach for testing and certification. Some have argued that examinations test competence; whereas performance can only be measured by observation (van de Vleuten & Schuwirth, 2005). However, for examining the effectiveness of experiential learning modes for the purpose of educational evaluation, the usefulness of subjective assessments of competence lies in the insights gained from groups of trainees. This information might be viewed as providing a pathway towards understanding how demonstrated competence might be achieved.

Summary of the Origins of General Practice Vocational Training

This chapter provided a glimpse of the background to the development of the discipline of general practice, and the history of general practice training in Australia. This history included the movement from a relatively unrestricted program to a highly structured one. The chapter included some of the key pressure points in negotiations between the Royal Australian College of General Practitioners and the Department of Health. It also featured the work of Benner (1984), as a means with which to understand the development of general practice competencies. Finally, Miller's pyramid on the assessment of performance-based assessment (1990) was considered. Herein lies the link between performance-based assessment and self-assessment, which placed experiential learning methods at the forefront as a means with which to understand the pathway toward demonstrated competencies. Chapter three examines trends in education that were instrumental to general practice training in Australia by examining the field of self-assessment studies in general practice vocational training and the role and function of self-assessment.

Chapter 3: Perspectives of Learners

This chapter considers an established typology of self-assessment studies based on their different uses. Consistent with the diverse uses of self-assessment in these studies, a series of different themes emerge. Several concepts are discussed in terms of self-assessment. These include the distinction between self-assessed competence and confidence and the interdependence between the two concepts. What becomes evident, particularly in the section of the chapter on the current frameworks of self-assessment is that self-assessment in medical education is currently undergoing a process of consideration and re-evaluation. After considering these uses for self-assessment in medical education, a model for application to general practice training in Australia is highlighted. This model is developed further in the self-assessment study in Chapters five, six and seven. To establish the relationship between self-assessment and the enhancement of education methods, a distinction will be made between self-report and self-assessment.

The Field of Self-assessment Studies and General Practice Vocational Training

In two landmark articles, Gordon (1991; 1992) reviewed the validity and accuracy of the previous twenty years of self-assessment research in the education of health professionals. The work of Speechley, Dickie, Weston, and Orr (1993) and Speechley, Weston, Dickie, and Orr (1994) followed, as part of a renewed interest in the role of self-assessment in fostering lifelong learning. Gordon highlighted the notions of validity and accuracy of self-assessments, devising a typology of various genres of self-assessment studies that included knowledge testing, self-reflection, global assessments of performance, and innovative training programs.

Although Gordon (1991) found little association between student self-assessments and supervisors' judgements about their knowledge and skills, sometimes students assessed themselves more positively, and at other times, supervisors assessed their students more positively. This depended upon the context of the assessment. In the case of factual areas of the curriculum, for example, Gordon found low correlations between students and their supervisors. This is not surprising, for, as a rule, methods for validating competencies such as correlations have been shown to require a very high level of congruence indeed (Falchikov & Boud, 1989). Self-assessments, on the other hand, as subjective measures, do not necessarily measure the competence of the individual, but can be used to ascertain the responses of groups for educational purposes. It was in relation to an educational context that Gordon recommended self-assessment could improve in accuracy over time. A parallel situation existed in Australia with the use of self-assessment to examine the key influences on general practice vocational training.

It is of some note that some researchers considered that students were able to assess the skills of their peers with some accuracy, but were unable to assess themselves. This has some precedent in the literature as discussed by Arnold et al. (1985), Calhoun, Wooliscroft, Ten Haken, Love, and Boxer (1986) and Calhoun, Ten Haken, and Wooliscroft (1990). In contrast, Gordon (1991) depicted self-assessment (through the provision of clear goals and specific training) as a learnable skill. He extended this further and concluded that the technique of self-assessment was under-investigated because trainees were not in the habit of engaging in it. Gordon found data on self-assessment to be variable and difficult to compare across studies—a point also amplified in a subsequent meta-analysis by Falchikov and Boud (1989). This leads to the conclusion that because self-assessment is not widely practised, there is a need for self-assessment teaching and research in preparation for the autonomous life of professionals after training.

Gordon's review (1991) served to reawaken an interest in self-assessment as an essential, but under-developed skill for adult learners in autonomous professions. Both of Gordon's articles re-invigorated the discourse around self-assessment by providing a typology based upon the different uses of self-assessment as a useful and useable technique. It is timely to reconsider his typology specifically for medical education, and to ascertain whether Gordon's initial observations on the lack of valid self-assessment studies and the promise of longitudinal evaluation are still applicable to the field of general practice vocational training.

A current search of the Medline and the Educational Resources Information Center (ERIC) databases for the keywords of self-assessment, evaluation, monitoring and report yielded 35 studies relevant to medical vocational training. All of the studies were from 1970—2006. Of 44 studies relevant to the overall field of self-report or self-assessment, 15 related to the educational evaluation, 12 focused on the validity of self-assessment, two were meta-analyses and six examined the extent to which self-assessment might be a learnable skill. Nine further studies belong to a generation of self-assessment contributions that use current frameworks to understand the educational process. (See Appendix A 1.)

Overall, several key questions initiated by Gordon continued to engage the interests of researchers on self-assessment. These questions related to the contribution of self-assessment to educational evaluation, the extent to which self-assessment might be taught and the issues relating to its validity. However, the field of self-assessment also needed to consider and reflect a multiplicity of disciplines, ranging from education to social science. What follows in this chapter describes and appraises the genres of self-assessment for the purposes of evaluation, education, and validation to place the use of self-assessment in the study in Chapters five to seven into context.

Self-assessment and Educational Evaluation

With the establishment of general practice as a discipline, government, policy makers, and medical educators have become increasingly interested in the ability of medical educators to demonstrate the effectiveness of their educational programs. Broadly speaking, the evaluation of general practice vocational training literature on self-assessment has focused on attitudes of trainees to training and the extent to which training enhances knowledge. This is part of an approach to evaluation that considers the value and effectiveness of an educational or training program from the perspective of learners. Described here are the topic areas of this research into attitudes to training, the context of this interest, an exploration of what is being measured by competence and confidence questions, and the relationship between self-assessed competence and satisfaction, and the importance of sub-group analysis.

Attitudes to Training

When assessing the appropriateness and quality of general practice vocational training, the focus of research has been on trainees' attitudes to educational programs along with their own competence. The methods used in these studies range from satisfaction surveys, cohort studies, studies on global skills, and analyses of curricula that focus on strengths and weaknesses.

Medical education researchers have examined the attitudes and views of participants of educational programs in topic areas such as trainees' self-assessed competence (Arnold et al., 1985; Goedhuys & Leroy, 1990) and their preparedness to practise or confidence and self-assessed competence (Biro et al., 1993; Donnelly et al., 1994). These terms have been used interchangeably. The quality of education and satisfaction with training have also been the subject of research enquiry (Bligh & Slade, 1996; Clack, 1994; Clack & Head, 1999; Girgis, Sanson-Fisher, & Walsh, 2001; Thomas & Snadden, 2000). Various

medical educators have drawn on the psychological sciences by focusing on the extent to which training facilitates desirable attitudes in learners (Grol, Tielens, & Mokkink, 1985; Speechley et al., 1993; Speechley et al., 1994). Generally, in this area, researchers have focused on the measurement of a single attitude rather than the application of a comprehensive educational theory. One explanation for this might relate to the challenging nature of applying theory to educational evaluation.

In the 1970s, however, medical educators began to embrace progressive developments in education, viewing the examination of a fixed body of scientific content as insufficient for the assessment of learners. Instead, medical educators began to consider the importance of research examining the attitudes of trainees as an influence upon behaviour. The work of Kegel-Flom (1975) was an early example of this focus on the attitudes of trainees. In her view, “The performance of medical interns, like that of physicians in general, is apparently associated more with personal qualities and medical training than relative levels of aptitude and undergraduate achievement” (p. 815). At a time when the role of adult learning was gaining in significance, researchers described a more holistic form of assessment than their predecessors did, by bringing the individual qualities of the trainee to the foreground in the learning process.

These developments on research into attitudes to training were evident in a body of work in North America, and generated new perspectives on the significance of self-assessment. In the evaluation of family medicine programs in the 1980s, Arnold et al. (1985) examined the self-assessed competence in internal medicine of baccalaureate medical students at the University of Missouri, Kansas City during a four-year longitudinal study. Although not solely designed for the purpose of evaluation, their study can be viewed from this perspective in an effort to establish a relationship between self-assessed competence and satisfaction.

In this endeavour, Arnold et al. incorporated multiple perspectives by seeking to explain the reasons for the differences between trainee and supervisor ratings observed in the preceding literature (Keck & Arnold, 1979; Kegel-Flom, 1975; Morton & Macbeth, 1977; Willoughby, Arnold, & Calkins, 1981).

Arnold et al. (1985), in establishing modest correlations between supervisor and trainee ratings, concluded that supervisors and trainees measured different concepts in their assessments. Rather than valuing one perspective over the other, they recommended that there was scope for the use of either form of data—particularly for educational purposes.

Supervisors' assessments, on one hand, provided an assessment of the capabilities of the trainees, more appropriately measured through criterion measures of cognitive achievement. Trainees' assessments, on the other hand, were, by their very nature, subjective. To extend this idea further, the views of trainees can be said to offer insights into the educational appropriateness of a program. In this context, it is as if groups of trainees are answering the following question: "Overall, in which areas of the curriculum do you see yourself as being strongest and weakest?" The important point that arises from this is that if the two groups are measuring different things, using different forms of data, the results are unlikely to be consistent. Such discrepancies, however, do not render the use of such data as invalid; but are dependent on the context in which they are used. For the purpose of understanding educational processes, the absolute values so important to summative assessment are neither appropriate nor practical in understanding how findings on programs are derived. It is this use of self-assessment for educational processes that extends the findings from Arnold et al. to the model at the core of this dissertation in Figure 1 on page 94.

Another key point arising from a discussion of the longitudinal aspects of the work of Arnold et al. (1985) concerns whether self-assessment might measure what it is claimed to, or the validity of self-assessment. The increase in self-assessed ratings suggested their validity because the ratings might be expected to increase. The validity of measures derived through self-assessment was also supported in other literature (Arnold et al., 1984; Falchikov & Boud, 1989). Moreover, the use of self-assessed competence ratings in an evaluation context, applies to group findings, not individual summative assessments for certifying general practitioners as safe to practice. This distinction to make in the use of self-assessment measures is an important one.

The use of longitudinal research by Arnold et al. (1985) found further resonance in the literature in support of the use of self-assessment in the area of health research; but this differs from discipline to discipline. Although Kruger and Dunning (1999) in the field of psychology have questioned the ability of individuals to judge their own abilities, this research pertains primarily to the identification of a lack of competence. Researchers in behavioural sciences, on the other hand, accept the use of self-report data in the field for the assessment of physical activity (Glasgow, Ory, Klesges et al., 2005). The use of similar methods to ascertain smoking prevalence also relies on the fact that by comparing populations from year to year, the question of over- or under-reporting becomes redundant. In the case of self-assessed competence, it is reasonable that individual respondents might well over- or under-report their levels of competence; however, as a group, this should occur with some consistency from year to year. Further, the issue of absolute values becomes less important when findings are compared across time. Although Gordon (1991) placed some emphasis upon longitudinal improvement as evidence of program effectiveness, particularly in the context of educational evaluation, the importance of this idea has not received the acknowledgement that it might have deserved in subsequent research.

With the exception of Speechley et al., (1993) who examined the extent to which training facilitates confidence, relatively few researchers have considered the potential of longitudinal research in the evaluation of general practice vocational training. Possibly, this was due to the lack of comprehensive evaluation programs and outcome studies implemented on a long-term basis, as McClelland and Herlihy (1995) have observed of medical specialties in the United States.

Several points of relevance to the educational evaluation of general practice vocational training arise—most significantly from the work of Arnold et al. (1985). The use of self-assessment to derive attitudinal data was an acknowledgement of the changing nature of medical vocational training. The use of self-assessment as significant in educational evaluation enabled the group data arising from learners focusing on their own strengths and weaknesses to contribute to educational efforts. In addition, the validity of longitudinal designs from the behavioural sciences provided another avenue for the use of self-assessment—particularly when combined with the concept of multiple perspectives. Chapter five considers these aspects in a model designed to test the utility of self-assessment for evaluation purposes.

Self-confidence as a concept to evaluate the value of general practice training

Medical educators have frequently drawn on the term *self-confidence* (as distinct from competence) to develop measures with which to evaluate aspects of general practice vocational training. In this context, they have interpreted self-confidence to be a form of self-efficacy. It is appropriate here to provide a brief overview of discussion surrounding the meaning of this concept as a means to evaluate aspects of general practice vocational training.

Originating from the field of psychology, self-confidence was described as an individual's belief in his or her ability to accomplish a task (Bandura, 1982; Prochaska & di Clemente, 1983). This belief was considered to be influenced

by previous history, expectations of the individual about the difficulty of the task, and expectations of teachers and peers (Biggs & Moore, 1993). Educators have generally assumed that as students progress through training, so their self-confidence should increase.

One way of measuring self-confidence is by asking individuals about their competence in a given area. This relationship between the two concepts has fostered some confusion. Stewart et al. (2000) commented upon the significance of both self-assessed competence and confidence measures in describing the transition from novice to professional in medical education. The concept of competence was understood to be what a trainee knew about a task. This is distinct from confidence, which determines whether a trainee might undertake the task (Rethans, Norcini, Barón-Maldonado et al., 2002). Using this delineation, a high level of confidence was understood as perceived competence; whereas a low level was understood to represent anxiety. This meant that confidence did not present a linear construct. Instead, there was a sense of mutual interdependence between competence and confidence. For this reason, Stewart et al. (2000) have maintained the use of the concept *competence*, rather than *confidence* in the wording of questions to respondents.

Efforts to link self-confidence to an overall psychological theory of self-efficacy can be compared to the approach taken by Stewart et al. (2000). In a randomised controlled trial, Smith et al. (1995) proposed that confidence was a reliable predictor of the use of new skills in an educational intervention to enhance the psychological skills of medical students. If a sense of mastery was based on practice, this suggests a linkage between attitudes and behaviour provided that items measuring self-efficacy are related to specific stages as opposed to global mastery (Bandura, 1982; Prochaska & di Clemente, 1983). An acceptance of this

linkage between attitudes and behaviour was also evident in the early evaluation of general practice vocational training in Australia (Docker, Carson, Hughes et al., 1993).

An interrelationship emerges from the examination of the concepts of confidence and self-assessed competence. Donnelly et al. (1994) and Speechley et al. (1993) accepted the link between attitudes and behaviour, and focused on the utility of self-confidence as a measure of trainee comfort with their general practice skills. They regarded self-assessed comfort with general practice skills as a necessary prerequisite for competence, and found that trainees' comfort with general practice increased as training progressed. Speechley et al. (1993), derived similar findings and highlighted a novel application for self-evaluation: areas of strength and weakness might be used to identify trainees who need remediation. This built upon the work of Arnold et al. (1985) around the importance of an individual's strengths and weaknesses. The consideration of these group data places such a project into the realm of educational evaluation.

In general practice vocational training programs, the stated outcome is the ability to practise as a competent, independent general practitioner. An examination certifies this ability. Facilitating the skills involved in independent practice might entail an understanding of the process of transition from hospital doctor to independent general practitioner. Examining the attitudes of the practitioners, particularly their self-confidence might inform this. Despite a possible link between the attitudes of practitioners and their later behaviour, the ideas presented by Stewart et al. (2000) have not been widely implemented. This might be because few programs in general practice vocational education have been subject to comprehensive evaluation. Although Australian general practice vocational training attempted to embrace self-confidence from the outset, it was not based on a deep understanding from the psychological sciences. Nevertheless,

the key message of these researchers on the value of self-confidence remains: identifying areas of strength and weakness can identify aspects of a program that program developers might improve.

The issue of the morale of trainees relates to general practice vocational training, because this, in turn, affects completion rates. In Australia, where general practitioner workforce distribution and shortages are a problem, the issue of completion rates is pertinent in a policymaking sense. In addition, confidence is significant for its effect on the willingness of general practitioners to engage in any given clinical or behavioural activity. This has relevance to the area of continuing medical education. In summary, several points that relate to both confidence and self-assessed competence in the evaluation of general practice vocational training have emerged. Some lack of clarity surrounds the use of the concepts. A satisfactory understanding of these terms is as mutually interdependent. The importance of them is in providing an understanding of learner morale and later clinical behaviour especially with regard to general practitioners' willingness to engage in desired practice activities. The study in Chapters five, six and seven examines further the phenomenon of self-assessed competence.

One approach to the evaluation of training draws on social science methods to examine the outcomes of different cohorts and to evaluate the value of the program by assessing whether it meets its objectives. In the case of medical educational evaluation, one of the stated outcomes of a program might be that of enhanced confidence or readiness for independent practice. Biro et al. (1993) compared self-assessments of family medicine graduates and specialists. They reported a positive relationship between the number of patients seen by students and their self-assessed competence. It might be hypothesised that the more supervised practice students have, the more their self-assessed competence improves.

Trainee doctors have also exhibited gender differences in self-assessments. At King's College School of Medicine in London, for example, women reported better listening skills, teamwork, compassion, and interpersonal skills. Men, on the other hand, reported being better equipped for leadership and the tolerance of ambiguity and uncertainty (Clack & Head, 1999). These findings are of significance to curriculum development and to professional life after training, as women are an increasing part of the general practice workforce in Australia.

Researchers compared the self-assessments of different types of practitioners with fruitful results. Girgis, Sanson-Fisher and Walsh (2001) compared the perspectives of physicians, surgeons and general practitioners on their own capabilities in a range of communication domains, and areas of postgraduate training. Their study provided the basis for recommendations about vocational training—particularly in areas pertaining to public health. Only a small percentage of general practice trainees assessed themselves to be competent in preventive care. Girgis, Sanson-Fisher and Walsh contributed to the points raised by Carlisle (2000), Epstein and Hundert (2002) and Stewart et al. (2000). Their work places the linkage between self-assessed competence and actual performance into some context: low levels of self-assessed competence might be mirrored in the avoidance of treating particular conditions in prescribed guidelines. This further supports the use of self-assessments in vocational training program evaluation, and suggests that there is considerable scope for further research into self-assessment and adherence to guidelines in light of public health objectives. The literature on sub-group differences suggests that examining group differences is an accepted method for evaluating training. This is particularly the case in situations in which training has become the only pathway to vocational registration, and control groups for randomised controlled trials no longer exist.

A series of studies have examined self-assessed competence in the evaluation of educational interventions (Biro et al., 1993; Bligh, 1992; Clack, 1994; Donnelly et al., 1994; Goedhuys & Leroy, 1990; Grol et al., 1985; Speechley et al., 1993). Trainees' satisfaction with their learning experiences has also been used to provide an understanding of how to enhance the curriculum, and training experiences more generally. In some instances, satisfaction with training was examined by considering the extent to which participants assign a high or low rating to their training experiences (Clack, 1994; Hojat, Gonnella, Veloski, & Erdmann, 1996). In other studies, satisfaction with general practice, mobility, gender differences and career paths has been used to develop recommendations about changes to general practice vocational training in Britain (Bradley, Murphy, & Lambe, 1996). A central idea emerging from these studies is the relationship between self-assessed competence and satisfaction. In this context, satisfaction is important to initiatives that support self-directed learning as a central aspect of general practice vocational training.

Summary of Self-assessment and Educational Evaluation

This examination of the application of self-assessment to evaluation underscores the importance of Gordon's contributions (1991; 1992) in reinvigorating the field of self-assessment in the health professions. Researchers in the field of self-assessment and education have built upon Gordon's study categories. The most relevant contribution arising from this work has been that of Arnold et al. (1985). In this context, self-assessment became analogous to answering the question of which areas of the curriculum learners see themselves as being strongest and weakest. In this context, Arnold et al. asserted that self-assessments are to be delineated from supervisors' assessments of their students' capabilities. This serves to remind us that self-assessment, as an educational and evaluative tool, is not to be confused with the certification of a general practitioner as competent.

The other contribution of Arnold et al. (1985) is in their advocacy of the application of longitudinal evaluation designs. With the increasing specialization of medicine, training is becoming mandatory, and evaluators, in many countries find that randomised controlled trials are no longer feasible. Under such circumstances, some researchers represent the longitudinal evaluation design as a valid alternative in educational evaluation.

The approach used by many researchers examining self-assessment owes much to social science methods, particularly the approach of sub-group analyses. For example, researchers accept the improvement in trainees' self-assessed competence as learning progresses as indicative of the appropriateness of training. This is because for educators, the purpose of indicators of appropriateness particularly relates to educational processes. Of the evaluation studies of educational programs that focus on attitudes to training, little of the research has an articulated theoretical basis in terms of an overriding psychological theory (Falchikov & Boud, 1989). Despite this, some studies in this area might potentially contribute to the improvement of educational programs. The relationship between satisfaction and self-assessments is particularly noteworthy in this regard.

Several implications arose from this examination of self-assessment and evaluation. The public health agenda has competing priorities. The educational evaluation studies examined suggest that programs seeking to encourage particular behaviours in general practitioners need to be designed with the attitudes of general practitioners in mind. Although policymakers and public health practitioners might prefer that general practitioners addressed many issues, this is not always possible within the constraints of the general practice consultation. The relationship between effective education and improved patient care is somewhat indirect, but it does follow that low levels of self-assessed competence might lead to practitioners who avoid treating particular patient

problems and presentations. This is the rationale for closely examining the attitudes of general practitioners during and after their training as a method for evaluating programs.

Validation Studies

Self-assessment studies examining the effectiveness of educational programs by measuring whether programs facilitate enhanced knowledge and skills examine multiple data sources to see whether there is a correspondence between them. In these studies, the knowledge and skills of medical students have been defined in relation to the areas of basic science, clinical knowledge, management and history. This range of topic areas, all considered integral to family medicine, bears reasonable similarity across programs of the United States, the United Kingdom and Australia.

After self-assessment for educational evaluation purposes, the validation of self-assessment with other forms of assessment is the next most popular research pre-occupation. Fourteen out of the 36 self-assessment studies in Appendix A 1 sought to explore the validity of self-assessment in medical education faculties. The studies were particularly noteworthy for the questions they raised about the comprehensiveness of the measures compared in the validation studies. Antonelli (1997), Fincher, Lewis, and Kuske (1993), Keck and Arnold (1979), Kegel-Flom (1975), Kolm and Verhulst (1987), Linn, Arostegui, and Zeppa (1975), Risucci, Tortolani, and Ward (1989), Sciabassi and Woelfel (1984), and Wooliscroft, Tenhaken, and Smith (1993) studied faculties in North America. Fraser, McKinley, and Mulholland (1994) examined faculties in the United Kingdom. Jansen et al., and Jansen (1998; 1995) worked in the Netherlands. The work of Farnill, Hayes, and Todisco (1997) was from Australia and that of Morton and Macbeth (1977) was from New Zealand. The studies focused on skills, and non-

cognitive measures as well as knowledge. Measures examined in these studies bear some similarity across the countries of the United States, United Kingdom, the Netherlands, Australia and New Zealand.

The focus on the validation of self-assessment with other forms of assessment is understandable because the value to educational researchers, were self-assessment valid would be considerable. Primarily, the value of self-assessment relates to the development of autonomous professionals and the acknowledgement of limitations. In this regard, the role of self-reflection is crucial. It seems intuitively correct to establish the meaning and significance of self-assessment by comparing it to other forms of assessment. Unfortunately, the appeal of the validation of self-assessment with other data contributes to a situation in which correspondence between data sources has been viewed as being synonymous with accuracy (Falchikov & Boud, 1989; Harris & Schaubroeck, 1988). Consequently, the search to validate self-assessment data continues while questions arise from the literature as to why this approach should necessarily continue.

Questions Around Validation Studies and Self-assessment

Researchers, especially in the United States, have taken a particular interest in the variability of ratings between supervisors and students. Despite a body of literature suggesting otherwise, they have maintained the expectation that the assessments of trainers and trainees should be consistent. In many studies, researchers have found that trainees rate their competencies lower than do their supervisors. From this, some researchers have concluded that supervisors and trainees represent different perspectives on assessments. They have concluded that supervisors have taken a more global approach than do their trainees who have been more sensitive to the differences between various topic areas. Kolm and Verhulst (1987) found very different factor structures of ratings in comparing self- and supervisor ratings. For supervisors, professional behaviour and clinical

competence accounted for most of the variance in competence outcomes, but for trainees, professional behaviour, interpersonal skills and relationships accounted for most of the variance.

These findings suggest that the conceptual basis of research that seeks to validate one form of data with another does not necessarily suit the needs of educational research. For example, it might not be appropriate to benchmark the ratings of experts against other forms of data.

As Ward, Gruppen and Regehr (2002) have suggested, “There is good reason to believe that experts are not measuring what they intend to measure, or that what they intend to measure is necessarily important” (p. 69). Their comments not only address the issue of whether different data sources will necessarily correspond, but also question the status and relevance of what the expert might have to offer educational research.

Reliability is another methodological issue featuring in investigations of self-assessment and validation of other sources of data. In this context, reliability is defined as the consistency of answers from one person on multiple occasions (de Vaus, 2002). Falchikov and Boud (1989), in a meta-analysis, have cautioned against high expectations of reliability of self-assessments. This is particularly relevant to an educational context: “Given that experienced teachers are not reliable markers in all situations, then it is perhaps unreasonable to expect inexperienced students to always demonstrate reliability” (p. 427). Instead of the expectation that the views of self-assessors should be consistent with those of their assessors, this comment suggests that inter-rater reliability is even difficult within groups of the same types of raters. As an alternative, Falchikov and Boud have advocated a more equal, collaborative relationship between the observer who assesses the students, and the student who participates in the assessment process.

Such sentiments are consistent with the objective of enhancing the educational process. Kegel-Flom (1975) had already laid the groundwork for this approach. In reasoning that supervisors, peers, and medical trainees view performance in very different ways, Kegel-Flom challenged researchers to consider multiple perspectives when evaluating the work of teachers and learners.

This challenge informed the broader, interdisciplinary approach to general practice vocational training frequently reiterated by other researchers. With the exception of Arnold et al. (1985), Gordon (1991) and Regehr, Hodges, Tiberius, and Lofchy (1996), the priority for summative assessment and certification of general practitioners have overridden efforts to implement multiple perspectives of trainee performance, and medical education has not implemented this approach as energetically as it might.

Researchers conducting validation studies have made considerable efforts to ascertain the uses of self-assessment. To focus on the non-cognitive aspects of self-assessments, Morton and Macbeth (1977) compared the mean scores of staff, fourth-year surgery students and their peers. They found that after initial disorientation, participants became more comfortable after learning the skills of how to assess themselves. If a high level of confidence represents perceived competence, and a low level represents anxiety (Stewart et al., 2000), self-assessment might be used to identify students who require counselling and remediation during training. In this case, the focus of teaching self-assessment might not only be preventive, but also augment the central findings from Arnold et al. (1985) on non-cognitive or attitudinal variables to answer the question of an individual's perceived strengths and weaknesses. It also supports the position of Falchikov and Boud (1989) on a more collaborative approach to assessment.

Interpretations Around Difficulties of Validation Studies

Self-assessment validation studies have considered the question of whether self-assessment is a learnable skill in the range of areas synonymous with general practice skills in the United States, the United Kingdom, and Australia. One of the reasons offered for the achievement of less than desired correspondence between data sources is that trainees are unused to evaluating their own performance. With teaching, researchers suggest, this situation might improve. For Fincher et al. (1993), self-assessment is a learnable skill. They established that interns rated themselves as less capable than program directors overall (based on grade scores) on several dimensions of family medicine, including basic science, clinical knowledge, history taking, physical examination, ordering of tests, and appropriate management and referral.

Fincher et al. (1993) and later, Jansen et al. (1998) have presented arguments on the lack of congruence between supervisors and trainees which oscillate around the relevance of the curriculum and the measurement difficulties inherent in achieving correspondence between different data sources. Linn et al. (1975) previously offered a similar explanation by suggesting that ratings were a product of the environment. Each of the studies highlight the value of testing both an intervention for the teaching of self-assessment, and the capability of adult learners to assess their own strengths and weaknesses. This implies that self-assessment might be appropriate when applied to the purpose of evaluating curricula.

Of those studies located in the area of self-assessment, only two meta-analyses have some application to general practice. The findings of a substantial review concerned the question of self-assessment compared to other data sources. Harris and Schaubroeck (1988) adopted a psychological approach analysing the

comparisons between self, peer and supervisor data from 64 studies. Falchikov and Boud (1989), on the other hand, in an analysis of 57 studies used a more educational perspective. They found that self-assessment was a teachable skill, and trainee and supervisor ratings became closer as time progressed. They also found more consistency in the results of trainees and their supervisors in scientific disciplines. This extensive study suggested that validation was much less of a consideration in the context of educational evaluation for the absolute precision necessary for summative assessment was not required to explore insights into the educational process.

Researchers have long noted the discrepancies between students' self-assessments and the assessments of their instructors. Sciabassi and Woelfel (1984) drew attention to the method being used to establish congruence as insufficient. Falchikov and Boud (1989) made a similar point by demonstrating just how much congruence would be required for trainee and supervisor ratings to be in agreement using the correlation method. Ratings would need to be almost identical to produce a reasonable correlation. This meant that the benchmarks for achieving validity were unusually high.

Further self-assessment research uses different methods for exploring consistency between the self-assessment ratings of supervisors and trainees. Keck and Arnold (1979) have found that attitudinal or non-cognitive variables provide better predictions about how students might perform in examinations. Their argument about self-assessment is an interesting one: namely, that family medicine trainees are intellectually capable to the point where diminishing returns have been reached. If this argument applies, the measurable improvement in cognitive outcomes is minimal. It has been acknowledged that the grades of medical students are less able to predict their performance than their interpersonal skills in communicating with patients (Kegel-Flom, 1975). This moves debate on self-assessment into a different sphere. Instead of self-assessment being a tool with

which to testify to a trainee's competence, self-assessment becomes a diagnostic tool with which to examine students' strengths and weaknesses and make changes to the educational process. This makes a clear separation between self-assessment as an educational tool and the need to assure safe practitioners. The implication of these findings is that self-assessment might be appropriate for educational evaluation purposes when the method relates to the purpose of the assessment.

Summary of Validation Studies

Initial contributions on the validation of self-assessment research have focused upon the congruity of the assessments of faculty member, supervisor and student self-assessments. Notwithstanding the number of contributions taking this approach, the studies examined suggest that the methods used to establish congruity between trainees' and supervisor ratings are problematic.

As an alternative, some broadening out of thinking on the nature of validation methods arises from these studies. The key underlying assumption of many of these studies is that accuracy is the same thing as the correspondence of findings from different data sources. Accuracy is an issue when one is in the business of protecting public safety through the certification of general practitioners. The matter of assessing the value of an education program from the perspective of learners, on the other hand, warrants consideration of a broader array of approaches aimed at identifying their strengths and weaknesses.

The validation of self-assessment produced many different ways of using self-assessment. An important point emerging is that self-assessment, although not appropriate as a tool for certifying competence, is highly useful for educational evaluation, diagnosing where programs are functioning well, for testing how students respond to feedback about their performance, and for initiating the conversation about performance and the learning of the skill of self-assessment.

In summary, validation studies on self-assessment have shown that there is significant interest in the area. These kinds of studies are less useful as an end in themselves, and more useful for what they contribute to the process of education and evaluation.

Interventions to Teach Self-assessment

Calhoun et al. (1990), Cochran and Spears (1980), Henbest and Fehrsen (1985), Jansen et al. (1998) and Martin, Regehr, Hodges, and McNaughton (1998) are examples of researchers implementing interventions for students to evaluate their own performance against external criteria. An early study in this area from Cochran and Spears (1980) found that the ability of students to assess themselves improved as they progressed through a program designed to teach self-assessment. The notion that arises from this study is that self-assessment might be effectively taught if only the correct method were identified. Some time later, Henbest and Fehrsen (1985) found that after gaining familiarity with the assessment process and being active contributors to the development of the assessment criteria, there were correlations between the views of students and those of faculty. Farnill et al. (1997), in another study, have suggested that students were able to assess their interview skills to a level of accuracy comparable with other studies and in a manner that might be widely implemented. An explanation for these results is that students viewed the assessment process as a positive experience, and the culture around assessment was one of active involvement. Overall, however, these studies suggest that interventions to teach self-assessment use clear performance criteria. Despite the fact that clear performance criteria might not always be provided in educational evaluation, self-assessment might be said to offer some potential when the focus is on student perspectives of their education.

Self-assessment has been found to need reinforcement if the skill is to remain over time. Jansen et al. (1998) found in a randomised controlled trial, that it was possible to teach self-assessment techniques, but the results diminished after six months. Martin et al. (1998) also found that video feedback could be used to improve the self-assessment capabilities of residents because the examination of four video vignettes provided a set of agreed criteria with which students could evaluate their practice. However, similar to Calhoun et al. (1990), they concluded that although the self-assessment was conceived in a way that often led trainees to underestimate their skills, it still remained a learnable skill.

Summary of Interventions to Teach Self-assessment

There have been few attempts to actually teach self-assessment, and measure its effectiveness using agreed benchmarks. The seven studies reviewed recommended that the ability to assess oneself improves after learning the skill. Gordon's initial proposition that self-assessment can be taught remains pertinent. In addition, the provision of a collaborative learning environment contributes to the view of assessment as a positive experience.

Current Frameworks for Self-assessment

A debate around the need for rigour in assessment (as opposed to relevance in curriculum) continues in medical education, and is evident in the discourse around self-assessment. In relation to the terminology used, for example, studies focusing on self-report and self-assessment demonstrate subtle differences in the way the terms have been conceptualised. This has been particularly apparent in relation to current and emerging frameworks in the area. The differences become evident in varying perspectives on self-report and self-assessment based on their distinct purposes. Self-report, for example, has the connotation of reporting to another. In an educational setting, this sounds like an outcome. Self-assessment, on the other hand, connotes an educational process, whereby trainees assess

their own strengths and weaknesses to gain feedback and remediation. More current frameworks relate self-assessment for educational evaluation purposes to understand educational processes, as opposed to self-report, which is used for gaining data on practice behaviours.

The Need for Self-assessment in Medical Education

The need for learners' feedback on the learning process and the curriculum is a valid reason to foster the use of self-assessment in postgraduate medical education. As early as 1983, Ende (1983) reported on the phenomenon of vanishing feedback whereby supervisors in observing their trainees less frequently provided less feedback on their performance over time. Despite general agreement that supervisors should provide their students with feedback on performance, research on educational activities suggested that trainees continue to want more feedback than they currently receive (Cooper et al., 2001a; Cooper, Goldman, Jasper, & Wellard, 2001b; Cooper et al., 2001c).

One of the barriers to medical educators providing more feedback to students is the view that the process might compromise the summative examination. Gordon (1997) considered this issue and devised ways that medical education programs might provide more feedback to their students as they progress. He proposed the provision of a clear separation between formative and summative assessment that would avoid compromising the summative examination.

Hays and Wellard (1998) took a slightly different view on the implementation of in-training assessment, and were less in support of too rigid a separation between formative assessment and the summative examination. The in-training assessment approach potentially offers trainees reassurance during their learning and this might play a role in guiding curriculum design. More recently, Ginsburg et al. (2000) have investigated the use of self-assessment for the purpose of understanding developing professionals. An important aspect of this work is that

it examined aggregate results rather than the competence of the individual. From this, it can be inferred that self-assessments in the area of professionalism might be used to examine educational processes. Ley (1991) has also shown that information gleaned from similar group observation can be communicated back to curriculum developers to assist in refining training programs.

Benefits of Self-assessment

The work of Ginsburg et al. (2000), Regehr et al.(1996), Reiter (2002), and Stewart et al. (2000) is representative of some of the current frameworks highlighting the benefits of self-assessment, particularly in the preparation of autonomous professionals after training. Ginsburg et al. (2000) have drawn attention to the difficulty of evaluating the behaviour of developing professionals. Self-assessment might act as a useful starting point for a dialogue around professional lapses. When combined with feedback, it offers the potential educational benefits of a collaborative relationship in the context of the developing professional. Reiter (2002) supposed that the ability of trainees to evaluate accurately what they are capable of doing would make them potentially more secure learners, and ultimately safer practitioners. Stewart et al. (2000) in relating positive experiences of confidence to competence, and negative experiences to anxiety, made a relevant commentary on life after training, particularly in relation to what students choose to engage in during further medical education. Gordon et al. (1992) have examined the progress of the adoption of the initial recommendations of Gordon's previous study in a large number of studies. They found that educational self-assessment programs resulted in reduced stress, improved communication and self-directed learning, better motivation, and fewer errors. Researchers who seek to re-conceptualise the field of self-assessment attribute these benefits to the collaborative process inherent in the teaching of self-assessment.

Assessment as a Collaborative Process

Gordon (1997) advocated the separation between licensing and formative assessment, because in order to function well, self-assessment requires the involvement of the learner. He also concluded that this was unlikely to occur if the student focused solely on summative assessment. At first, students have found the self-assessment experience somewhat daunting. After a time, however, as the process becomes more collaborative, the experience becomes positive. Despite this, many medical faculties have been slow to adopt a model whereby students can engage in a dialogue and gain feedback after self-assessment. The implication of what Gordon (1991; 1992) described in his initial articles on self-assessment is far-reaching and consistent with principles of autonomy inherent in the practice of learner involvement. At the same time, however, Gordon distanced himself from some of the educational movements of the seventies:

Self-assessment does not imply data-free introspection. It implies a shift in responsibility for collecting, analysing, and interpreting credible evaluation data, and for deciding appropriate next steps in light of that information (p. 877).

In this description of the process of self-assessment, Gordon envisaged that students could take a lead role in identifying (with some precision) what they need to learn and how they need to learn it. This recommendation changed the emphasis of self-assessment away from validation. It advocated the use of subjective measures to examine the effectiveness and appropriateness of education programs, but with some element of empirical rigour. Westberg and Jason (1994) provided practical information on how Gordon's vision might be implemented. They recommended clear goals, standards, and outcomes, supervisor role models, and feedback, in addition to the participation of students in setting assessment criteria to enhance their reflective capacities.

The implication of Westberg and Jason's comments is that by following these recommendations with adult learners who will not suffer as a result, assessment can be a collaborative process, and can be the starting point for a stimulating and involving educational environment.

Measurement Rigour versus Educational Relevance

When Gordon (1991) conducted his study on the validity and accuracy of self-assessments in the health professions, he identified some of the shortcomings of previous self-assessment studies. In doing so, he provided signposts for future researchers on how to enhance the rigour of self-assessment studies. This resulted in a substantial body of studies validating self-assessments against external criteria. Gordon also recommended researchers to focus on methodology more broadly, and consider longitudinal studies in preference to singular time-specific studies. If few researchers adopted this strategy, it was presumably because of the relative difficulty of establishing extensive, long-term evaluation programs.

Efforts to evaluate programs have sometimes floundered due to the tension between the desire for research design rigour, and the need for educational relevance. Medical education, in particular, has suffered from this phenomenon, and yet, support for pragmatic designs came from a surprising quarter. Feinstein (1996) viewed as a father of the randomised controlled trial, showed considerable latitude in the choice of research design, suggesting that rather than focusing on a plethora of different forms of validity, it was sufficient for research designs to be relevant and appropriate (1996, p. 207). Feinstein's position resonated not only with research practitioners concerned with methodology, but also with medical educators grappling with the issue of how much rigour was sufficient to evaluate the progress of an education program. Feinstein also recognized the difficulty of moving from the randomised controlled trial to a naturalistic setting. For this reason, he advised that an armoury of validity testing should not deter the conduct

of research of potential value to health professionals and patients. In addition, he supported the use and application of psychometric approaches.

On the educational side of the spectrum, medical education and evaluation specialists, such as Prideaux et al. (1999) recommended that outcomes evaluation studies should focus on relevant and significant outcomes such as communication skills, satisfaction with learning, and lifelong learning, and the impacts of general practice education on the health system.

Inflexibility in the choice of research design can be a hindrance if it prevents program designers from evaluating their programs. This move toward design appropriateness from Feinstein has provided medical educators with the opportunity to renew their focus on educational relevance, particularly in the areas highlighted by Prideaux et al. This is, in effect, what the current frameworks potentially enable.

Changing Methodologies

If defining self-assessment is an educational practice in the pursuit of lifelong learning, in which individual students no longer judge themselves against their peers, a different method is required for testing its robustness. An extensive analysis by Ward et al. (2002) concluded that the attempt to validate self-report is likely to produce mixed results because studies can be subject to methodological problems. Whether comparing correlations or proportions of agreement between expert and self-ratings, each approach examined the individual's performance against a group. Ward posited a new paradigm around the use of self-assessment: namely that if the function of self-assessment is to show the learner their areas of strength and weakness, the point of comparison should be the individual not in relation to a peer group but in comparison with their own skills.

Summary of Current Frameworks for Self-assessment

Although the current frameworks for self-assessment are still evolving, they suggest several directions that take the debate beyond the tension between the need to certify general practitioners as competent, and the provision of feedback to students during training. For the task of educational evaluation or providing feedback, the focus might be on reporting group results of the descriptions of individual assessments of strengths and weaknesses. The results might contribute to an overall picture of the effectiveness of the vocational medical education program from the point-of-view of the learner.

For self-assessment, or indeed any kind of assessment to work effectively, the active participation of the learner is required, with students taking a lead role in determining what they learn and how. This does not mean that rigour in the choice of evaluation design should be abandoned but rather, as Prideaux et al. have recommended, it is possible to take the best of the biomedical model in testing the effectiveness of programs, and at the same time preserve the position of the learner at the centre of evaluation efforts (1999, p. 86).

Towards a Model for Evaluating Training

The Introduction featured a description of the Australian medical vocational training system. Under this system, trainee general practitioners participate in an apprenticeship model of training, and engage in everyday general practice under supervision. A qualitative analysis of the self-assessment studies reviewed in this chapter and Appendix A 1, particularly the work of Arnold et al. (1985) was the starting point for the development of a model to evaluate the value of training from the perspective of the learner as shown in Figure 1 on page 94.

Learners' perspectives of self-reflection are addressed in the part of the model depicting the relationship between self-assessed competence and satisfaction

using demographic variables such as the stage of training, gender, locality, as well as patient population and workload as the independent variables. Items of satisfaction with aspects of the program are intervening variables. Another intervening variable includes clinical feedback from supervisors, as an observable variable. Self-assessed competence and satisfaction with various aspects of training are the outcome variables. They serve to identify aspects of training that might explain the appropriateness of training for the learner and the encouragement of self-reflection and self-directed learning as objectives of training.

The model lends itself to the assessment of aspects of the program, by exploring what the outcomes are for learners, and how satisfied they are with their learning progression. The other aspects of the model, such as examination results, serve to augment trainees' perspectives. Further research after training can build upon the model, by examining the career paths taken by learners, and the effectiveness of the program from the perspective of their consultation patterns after training.

These aspects of the model and the story of its development are discussed in the self-assessment study in Chapters five to seven, and form the core of this dissertation.

Summary of the Field of Medical Vocational Training and Self-assessment Studies

This chapter has taken a descriptive approach to the literature on self-assessment as it pertains to medical education. Of all of the categories examined, educational evaluation emerged as the most relevant to the use of self-assessment as a basis for future learning and research. Notwithstanding this, studies validating self-assessments with other data sources were the most frequent. Frameworks that are more current also offer promise for the utilization of self-assessment in the study referred to in the Introduction.

The main points that emerged from an examination of 35 self-assessment studies were the importance of the longitudinal method in evaluation of general practice vocational training. In addition, the significance of self-assessment was noted because of the relationship between low levels of self-assessed competence and reluctance to manage particular patient presentations.

This chapter has investigated significant aspects of the literature on the direction of self-assessment for educational evaluation, validation studies, interventions to teach self-assessment, and the current frameworks around self-assessment. This analysis provides the framework for a study to examine general practice training from the perspective of the learner. Chapter four moves from the perspective of the learner to that of medical educators and supervisors.

Chapter 4: Perspectives of Medical Educators and Supervisors

As distinct from the emphasis of the previous chapter on the perspectives of learners, this chapter examines the perspectives of educators and supervisors in relation to educational theories to place the self-assessment study at the centre of this dissertation into context. The chapter begins with a brief description of several trends relevant to the incorporation of experiential learning as the mainstay of general practice vocational training. This leads towards an examination of ideas underpinning the teaching of the discipline of general practice. First, a presentation of contrasting philosophical and epistemological views provides a link to the role of experience in general practice vocational training. Second, a brief sociological description of medicine as characterized by uncertainty provides a link between experiential and lifelong learning and a corresponding need for individual practitioners to appraise their own strengths and weaknesses. Third, a critique around traditional learning modes provides a link to an understanding of the wish to enhance medical education and preserve the role of the family practitioner as healer. Self-assessment is presented as fundamental in addressing these needs. The explanation of the attachment to self-assessment within general practice vocational training provides some insights into the reasons for the support of the self-assessment project in Chapter five.

This chapter also provides an examination of the relationship between the educational theories of reflective practice and adult learning, current at the time of the development of Australian general practice training and the use of self-assessment. The model of experiential, practice-based learning, and Donald Schön's concept of reflection-in-action (1987), will be considered for

its contribution to the preparation of students for self-directed learning and its relationship to the model in Figure 1 on page 94 as the basis of the self-assessment study. An exploration of one of his case studies will demonstrate the application of this approach to general practice vocational training in Australia as evident in focus group research with supervisors, and highlight the role of the general practice supervisor as a coach to the learner trainee in enhancing trainees' self-assessment skills. It is this use of qualitative research in combination with the quantitative methods of the self-assessment study that informs the findings. Both methods complement one another.

The remainder of this chapter explores adult learning and its incorporation into the Royal Australian College of General Practitioners vocational training program curriculum. This exploration is based upon insights gained from qualitative research with supervisors. The purpose is to provide a way of understanding the development of self-assessment, and the emphasis upon it within the program. A discussion surrounding the importance of experiential learning, lifelong learning, and adult learning in general practice vocational training highlights some of the significant critiques and the information provides a context for the relationship between theory and practice. This discussion of trends in teaching provides a context for considering the perspectives of medical educators and supervisors on trainee self-assessment, and provides a background for the self-assessment study that forms the centre of this dissertation.

Origins of Experiential Learning and Self-assessment

Since the beginning of general practice vocational training in Australia, experiential learning, as a principle, has been at its core. While there is scant evidence of the founders' main pre-occupations, an examination of their legacy points to three educational trends that influenced the rise of experiential learning in medicine: contrasting views of knowledge; the recognition of medicine as a

discipline of uncertainty, and the prevailing debate about how medicine was (and could be) taught.

The derivation of the experiential learning movement is apparent in an ongoing dialogue about the nature and acquisition of knowledge. This dialogue has its origins in modern philosophy. While rationalists claim reason to be the sole source of knowledge, empiricists claim experience to be the source of knowledge (Blumenfeld, 1996). These fixed positions usually prescribe that either one or other epistemological position might prevail. One consequence of this is that educationalists have adopted a critique of rationalism (rational analytic reasoning or positivism) to argue the benefits of experiential learning. The medical educational literature provides many examples of this philosophical stance to argue the benefits of experiential learning (Eraut, 1994; Malinen, 2000; Maudsley & Strivens, 2000b; Schmidt, 2000; Shapiro & Talbot, 1991). Their critique is, on the face of it, compelling; however, on further analysis, the scientific method need not stand in opposition to experiential learning.

In putting the case for change, advocates for the reform of education, present a particular view of knowledge in contrast to the tradition of rational analytic reasoning. Dreyfus and Dreyfus (1986) proposed a model that distinguished between knowledge and understanding by presenting understanding as both an achievable skill, and a method for survival in the social world:

human understanding was a skill akin to knowing one's way about the world rather than knowing lots of facts and rules for relating them. Our basic understanding was thus a knowing *how* rather than a knowing *that* (p. 4).

This delineation between knowledge and understanding represents a possible pathway for the acceptance of experience as a route to the acquisition of

understanding, and the need for applied knowledge. The acknowledgement of experience as a valid form of learning can also serve to define learning as a social endeavour, as distinct from being one that necessarily occurs around the individual.

The debate around rationalism (rational analytic reasoning) and empiricism (positivism) also occurs in the sphere of adult education. Kolb (1984) for example, identified the significance of epistemology as the source of disagreements about how to implement educational programs for adults. His explanation of the topicality of the distinction between rational analytic reasoning and positivism refers to his own experience as a pedagogue. Kolb explained that he had, "... often seen barriers to communication and problem solving that at root are epistemologically based—that is, based on conflicting assumptions about knowledge and truth" (1984, p. 37). These barriers might underlie efforts to gain control over content and methods of teaching, particularly in professional disciplines. They have remained topical in a continuing discourse arguing the benefits of the scientific method above social science in the medical curriculum (van Weyden, 2004) and remain at the heart of the arguments of researchers who view that students are not able to assess their own skills (Liddell, 1999). While it is to be acknowledged that experience might be flawed, the skilled educator is able to teach students to appraise critically their own skills and to have some expertise about what emphasis to place on experience. This objective of professional self-assessment has always been embedded in the general practice vocational training curriculum in Australia.

Although both compelling, and continuously documented in education literature, the debate around knowledge and experience need not be as polarized as presented in some of the literature according to Bertrand Russell (1962). He views Kant as developing the compromise between rationalism and empiricism

to provide the milestone that enabled a move away from the polarization of this debate. As Kant indicated in *The Critique of Pure Reason*, the contribution of modern science poses the question of whether synthetic judgments (acquired only through experience) might have a basis other than experience. Kant concluded that it is our own thought processes that order matter in space and time, and provide the concepts with which to understand experience (Kant, 1781). This position represented a compromise between the position of all knowledge as derived from reason and the acceptance of the role of subjectivity, intuition and experience as valid forms of knowing. Because of this development, experiential learning became an integral feature of educational theory.

The compromise position on rationalism and empiricism makes a compelling case for valuing experience in the context of a multidisciplinary area such as adult learning. The value placed upon experience as a learning pathway, and the definition of learning as a social rather than an individual endeavour, has had a wide influence upon educationalists. The value of experience in learning without the ability to appraise one's strengths and weaknesses is potentially limiting. For this reason, the founders of Australian general practice vocational training, influenced by these directions in their advocacy of curriculum innovation in medical education, incorporated self-assessment skills from the beginning. The fostering of these skills is the basis of the self-assessment study in Chapter five.

Uncertainty in Medicine

Another trend, based on the uncertain nature of medicine, argued for the use of experiential learning based methods in order to foster an advocacy movement around lifelong learning in medical education. The origins of the linkage between this uncertainty thesis and medical education lie in sociological debates, and in particular around notions of power and authority, as against the need to preserve the ethos of the family medicine practitioner. The practice of medicine has been,

from the outset, one in which the more that is known the more uncertainties abound. Renée Fox (1957) has advocated over many decades for a sociological approach to the uncertainty thesis and limitations of current knowledge.

An important aspect of the uncertainty thesis (for medical education) is the view that learners need not only to identify and concede their mistakes, but also to learn from them. Nevertheless, this need to learn by doing (as supported by experiential learning advocates) has not been systematically addressed in the transition from novice to medical professional until relatively recently. With the exception of the implementation of problem-based learning, for example, curriculum innovations have not been universally implemented in undergraduate training in the United States, the United Kingdom and Australia (Dowton, 2005; Liddell, 1999). One reason offered for this reluctance to embrace change is that, in some respects, the uncertainty thesis remains problematic for some medical professionals. To concede a lack of knowledge represents a challenge to the control exerted by the people who are valued as professionals. Early in the history of sociology, Parsons (1939) identified the conflict between the healer and that of the medical care professional. In Parson's view, this conflict was, through his belief in the ethos of the medical care professional, a resolvable one.

Fox (1957), whose mentor was Parsons, advocated for more clarity in distinguishing the difference between incomplete mastery and limitations of current knowledge. She represented a key forerunner of the trend of uncertainty in medicine. With the use of participant observation of patient-centred research, Fox (1980) continued this sociological approach seeking to explain why, as time progresses, less is known about the complexities of health, illness, life and death. Her work opened the way for the recognition of the need to prepare learners not only for lifelong learning but also for the skills of assessing their own strengths and weaknesses.

With the rise of the consumer movement, sociologists developed medical dominance theory of the 1970s, and challenged Parson's benign perspective of the medical professional. Freidson (1970), and recently Waitzkin (1991) and Willis (1989), for example, offered a critique of the position of medical practitioners based upon the control they exert over other health professionals and patients. At the same time, some questioning has occurred of the pre-eminent position of family medicine practitioners in the schema of dominance theory (as described by Freidson). This has occurred at a time when medical practitioners in becoming increasingly specialised have faced deskilling and a corresponding reduction in status through the intrusion of bureaucratic structures into what was formerly a very autonomous environment (Haug, 1973). The challenges of uncertainty, and critique of medical dominance theory and increased specialisation, provide the context for medical educators in Australia. In response they have sought to position experiential modes of learning as a means to preserve the ethos of the family medicine practitioner—as healer, who learnt by doing, was able to learn from mistakes, understand individual limits and recognize the need for lifelong learning.

The Critique of Traditional Models of Learning and Teaching in Medical Education

The critique of traditional learning models in medical education has been largely due to the perception of a lack of learner control and involvement. The vision of a large number of students sitting in rows with their hands on their desks and visible to the teacher is often associated with the industrial model of the 19th century and is more reminiscent of the panopticon from Foucault (1979), than that which might be regarded as enlightened schooling. When applied to the field of professional education, this vision was analogous to the subject-centred learning model with little opportunity for students to determine either content or learning pace (Wellard & Edwards, 1999). Under such a system, it was up to

students to make the link between subject matter and professional practice. As such, the approach assumed little individuality with regard to the requirements of the learner. As a view of learning, this approach did not acknowledge the benefits of learner-centred education—a key tenet of the educational reforms of the 1970s and 1980s. Further, a critique of traditional learning models has addressed concepts from the psychological sciences, such as learner self-efficacy for their contribution to learning and pedagogy. Other skills, beyond the mechanism of mere memory, such as higher order thinking and meta-cognition, have achieved recognition as medical education policy makers have become more multidisciplinary in their approaches (Marienau, 1999). Despite calls for reform, academic medicine and medical education have been reasonably slow to incorporate some of the key changes to pedagogy called for in this body of educational literature from the 1970s and 1980s. On one hand, educators from this era sought to incorporate the concepts of the pre-eminence of the learning experience and learner control in an effort to promote lifelong learning. The advent of general practice vocational training in Australia offered the opportunity to engage also in learning as an idealistic endeavour. On the other hand, the nature of general practice as a specialty provided a tension between idealistic educational theories and the day-to-day realities of general practice.

Pragmatic matters, such as how to define and certify professional competence and predict practice behaviour as outcomes of training took priority as the program progressed. The founders also recognized the need to examine the circumstances around the acquisition of general practice competence and the process of learning. This provided the linkage between objective assessments and subjective self-assessments, and established the basis for the evaluation of the program.

The critique of traditional learning models, for their lack of emphasis upon the perspective of the learner, their philosophical underpinning preferring

factual knowledge above heuristic skills, and the lack of consensus about the demonstration of clinical competence occurred, therefore, within a particular social context. This context included a greater awareness about the level of uncertainty in medicine, and cognizance of the need for newer skills, such as meta-cognition and evidence-based medicine. These developments represent an acknowledgement by departments of academic medicine that it is becoming increasingly difficult to specify a finite body of knowledge for the lifetime of a practitioner. Instead, medical educators in the United Kingdom, Canada and Australia have articulated lifelong learning as an important tenet of their programs and with it recognition of the need for self-assessment. To some extent, general practice vocational training in Australia arose from a growing critique of traditional learning models and an increased awareness of influential educational theories.

Influential Theories in Australian General Practice Training

Discourse around the program of Australian general practice training has always highlighted that, as an objective of training, trainees should become competent, independent practitioners. Implicit to this objective, was a view that placed both competence and independence on a continuum, as opposed to representing a binary position of either the presence or lack of competence. Consequently, the founders of general practice vocational training linked the development of self-directed learning to the effectiveness of the program. It was for this reason that self-directed learning became established as a desired goal for the completion of the program (Prideaux, Lyons-Reid, & Ashenden, 1995). As the program developed to incorporate summative assessment, its focus became the certification of general practitioners and patient outcomes. A certain commitment to self-directed learning for adult learners as a key plank of autonomy and independence, however, remained within the curriculum and the system of

training as it evolved over the years. Part of self-directed learning requires some assessment of what one needs to learn, and recognition of professional strengths and limitations. It is in an effort to understand the commitment to self-directed learning, and examine the role of self-assessment in the study in Chapter five that some of the influential theories will be considered here.

As was already discussed in Chapter one, the development of general practice vocational training was influenced by the promotion of the discipline in Australia and overseas. Another approach to understanding the derivation of self-directed learning is to consider some of the theoretical influences affecting the program as it developed. These included the notions of experiential learning articulated by Agyris and Schön (1978), Kolb (1984), Knowles (1984) and educational practitioners and researchers who, in later years, began to influence developments in medical vocational training elsewhere (Eraut, 1994; Knowles, 1984; Malinen, 2000; Mezirow & Associates, 1990).

Ideas on the transition from novice to professional as developed by Schön (1987) and presented by Eraut (1994), Mitchell (2005) and Mezirow (1998) have been influential in the sphere of medical education. Knowles (1984) and Malinen (2000) have also focused on the considerations of Kolb on adult learning and types of learners. The work of Mezirow (1990) has also been continued by Candy (1991), Collard and Law (1989), Garrison (1997) and Malinen (2000). All of these theorists might figure in an explanation of the development of the educational models used in general practice vocational training in Australia. However, to provide a specific background to the study of self-assessment in the Royal Australian College of General Practitioners Training Program, the focus is on the selected works of Schön and Knowles.

Reflective Practice

The focus of Schön's work is in developing a model of experiential learning applicable to professional disciplines (such as medicine). It is in professional settings where precise answers cannot always be rehearsed in anticipation of real situations, and problems cannot always be solved through prior practice where creative thinking is required (Schön, 1987). These characteristics of professional practice, and the way that they relate to medical disciplines, have ensured that the work of Schön is closely aligned to the model developed for general practice vocational training in Australia. The following exploration of aspects of the work of Schön demonstrates its focus, theoretical contribution, areas of ambiguity and relationship to general practice vocational training as adopted in Australia with its emphasis on self-assessment. The exploration serves to link the perspective of medical educators and supervisors and trainee self-assessments with the theory of reflection-in-action as presented in Figure 1 on page 94 as the basis of the self-assessment study.

In *Educating the Reflective Practitioner*, Schön's observational project featured several case studies around the process of acquiring the skills of a professional. One particular case study demonstrated the value of experience in relation to different types of knowledge and skills acquisition in the field of architecture. Schön placed the learner at the centre of the interaction, at the same time preserved the role of the educator and expanded the teaching repertoire to include the coaching of the learner. The key feature of the case study between Petra, the architecture student, and Quist, the architect, is that Schön presented a theory about reflective practice by demonstrating the role of the facilitator.

The relevance of this case study is that educators have agreed that the essence of the demonstration was in the modelling of a design, provision of a theory, and feedback on the process (Eraut, 1995). Educators of professionals have therefore

adopted a broad set of strategies for practitioners working in professional roles under supervision, and this is evident in the general practice vocational training curriculum in Australia. An understanding of reflective practice has been assumed in professional education. Its exploration provides a means with which to understand the attachment to self-assessment within the general practice vocational training program. Also acknowledged is that observational research on reflective practice has remained elusive (Mitchell et al., 2005) and so an entry point into the concept of reflective practice is through an understanding of its theoretical underpinning.

Theoretical Basis for Reflective Practice

Schön (and other experiential learning proponents) emphasized the role of experience in becoming a competent professional by asserting that for the novice professional, the role of experience is paramount in making the transition to a competent practitioner (1987). Schön thought that direct contact through experience leads to tacit knowledge. Highly valued by educators, tacit knowledge is that which an experienced practitioner displays without conscious deliberation. In general practice, for example, tacit knowledge might be used in the process of identifying an illness, for example. This type of knowledge, described by Schön, is part of a framework that includes scientific knowledge (validated through scientific criteria), and stylistic conventions and professional standards (Eraut, 1994).

According to the theory of reflective practice, applied practice is the means by which to gain significant experience and develop the much prized tacit knowledge required for diagnosis and management, and patient-centred medical practice. This is particularly the case in Australian general practice vocational training in which patients in real settings present to trainees under supervision.

Stott and Davis (1979) developed a framework for the general practice consultation. When combined with the areas of the management of presenting and continuing problems, the opportunistic promotion of health, and the consideration of the causes of illness, it is conceivable that the broad strategies of modelling a design, constructing a hypothesis and providing feedback combined with the framework of the general practice consultation provide valuable signposts for the teaching of medical education and vocational training and the development of the reflective practitioner. The strategy of reflection-in-action described below outlines the relationship between general practice vocational training and the role of self-assessment.

Reflection-in-action

The following examination of the phraseology of reflection-in-action highlights some of its salient features in relation to the self-assessment study in Chapter five. The diverse interpretations of reflection-in-action produced a range of perspectives on the topic, with different approaches across different disciplines. Whereas educators more generally have explored reflection-in-action in relation to educational theory (Brookfield, 1995; Candy, 1991), medical educators have tended to draw upon reflection-in-action as a practical strategy with which to describe learning in medical practice (Campbell, Slotnick, & Parboosingh, 1999; Gruppen, White, Fitzgerald, Grum, & Wooliscroft, 2000). While there are a few exceptions to this approach, there has been no substantial observational research on either the process or outcomes of reflection-in-action (Mamede & Schmidt, 2004; Marienau, 1999).

In instances in which medical educators have explored theoretical concerns, they have predominantly applied the approach of cognitive theory to the concept of reflection-in-action (Malinen, 2000; Schmidt, Norman, & Boshuizen, 1990). They have tended to rely on framing theory to elucidate the concept. Derived

from Minsky (1977), a frame was described as a series of nodes. The transition from novice to expert in clinical medicine is said to revolve around learning about cases and the organization of content into nodes. This includes exposure to patients, development of an illness script for the emergence of contextual factors, and rapid decision-making when networks are transformed into succinct ones. The assumption has been that professionals are required to build a store of frames to the point where they might distinguish between a usual situation and one requiring conscious deliberation. Using the concept of framing theory, Eraut (1994) presented the process of reflection-in-action as one in which the practitioner was removed from the mechanical, everyday mode, to a mindset in which the current phenomenon and the thoughts around it were questioned. The salient point about the cognitive approach to reflection-in-action is that it is also a *reframing* process and, in the case of clinical medicine, facilitated by a supervisor. The role of the supervisor as evident in general practice vocational training might be therefore regarded as highly significant in the application of a cognitive approach to the concept of reflection-in-action.

A practical approach to the concept of reflection-in-action is to view it as challenging the individual, and as a proactive element in the learning process of professional self-examination. The supervisor in the context, acts as a critical friend, encouraging learners to reframe a problem, to stop and think, and answer questions that might lead to alternative perspectives (Eraut, 1994). In time, this process would assist trainees to examine their own strengths and weaknesses. Examples of qualitative research literature have reported on this relationship between student and supervisor among family medicine practitioners. Focus groups of trainers and trainees in the Netherlands identified the importance of self-reflection for the trainee, and the ability of the supervisor to encourage the process. Rather than a long intervention to teach self-assessment skills, self-reflection is considered part of learning. Boendermaker and colleagues (2000),

for example, found that the competent family medicine supervisor possessed the skill, desire and courage to offer feedback to the trainee. Similarly, in general practice vocational training in Australia, supervisors have reported on a process resembling reflection-in-action (Goldman, 1998). This process also emphasized the skills of engaging in observation and feedback, to facilitate improvements during the period of general practice training. In one of the focus groups conducted in preparation for the self-assessment study in Chapter five, a supervisor provided the following statement about the significance of the influence of the general practice supervisor:

It's very much about communication in the full sense of the word. The whole beauty of these attempts of any kind of direct observation is how much you can influence effective learning there and then on the spot at that critical moment so you can see big lifts [improvements] happen in a very short period of time (p. 19).

In general practice vocational training in Australia a range of medical practitioners provide feedback to trainees. The mechanisms of feedback should be considered in the context of several different roles in general practice vocational training. The role of the external clinical teacher served to enhance trainee learning experiences by observing trainees' consultations, and providing an assessment of the consultation (Kefalas, Usherwood, & Knight, 2005). The training advisor also reviewed trainees' learning plans and progress by advising supervisors and reviewing trainee performance. Medical educators, usually general practitioners with educational experience worked in state offices of the Training Program to assist trainees in reaching their desired learning objectives.

Another study on self-directed learning in general practice vocational training in Australia emphasized the importance of observation and feedback (Prideaux

et al., 1995). In this instance, a training advisor recommended the practices of the external clinical teaching visitor, as a way of encouraging and facilitating reflection-in-action. This was a way of cajoling the trainee into seeking more information about an area of perceived weakness. At the same time, the comments highlighted the importance of the dynamic between the external clinical teaching visitor and the trainee in augmenting and contributing to the training experience and encouraging the trainee to recognize their own limitations:

The visitors are on the spot. [During] self-directed learning they say, “You seem to know little about psoriasis. Have you thought about learning more about psoriasis?” and the trainee can’t possibly say, “No, I don’t like psoriasis” (Prideaux et al., 1995, p. 33).

The value of external observation and feedback might also serve to facilitate and encourage reflection-in-action for trainees in preparing for the habit of autonomous practice and lifelong learning. The value of personal biography, in the context of a comprehensive theory is that by naming a practice, teachers break the cycle of familiarity. This effort seeks to establish the habit of critical self-reflection in the trainee, and the learning of self-assessment skills in preparation for professional practice. This is important to the linkage between self-assessment and the recognition of one’s own limitations.

The essence of reflection-in-action, therefore, is considered as a somewhat experimental process linked to meta-cognition in which the learner is encouraged to question their understanding of the particular phenomenon at hand. It might be inappropriate to use such a process for every clinical encounter, for part of what signifies a professional is an ability to prioritise cases (Benner, 1984). Reflection-in-action as an experimental process, however, enables the student to test understandings, recognizing limitations, in order to inform subsequent

action. Eraut (1994) stated, "Reflection is essentially a meta-cognitive process, the thinking about thinking which informs decisions about what to do and think next..." (p. 146). Part of anticipating what to do next involves an understanding of where one is currently and that means an honest appraisal of strengths and weaknesses. This thinking from Eraut is consistent with contemporary educational theorists. In a response to this, medical educators of recent times, such as Williams (1998), have sought to incorporate the concept of reflection-in-action into curriculum design, and to render the concept as part of clinical training. Eraut it can be said, supported the view that reflection-in-action makes an important contribution to meta-cognition, particularly with the notion of reframing the reflective conversation.

Similar moves have been attempted in Australia with reflection-in-action and general practice training (Fabb, Heffernan, Phillips, & Stone, 1976; Prideaux et al., 1995). Trainees engaging in action research also observed their supervisors in the act of helping them to reframe clinical problems (Chien & Fennessy, 2004). The role of the supervisor as a form of coach in general practice vocational training in Australia, is particularly apparent in the strategy of reflection-in-action. The process has been highlighted in efforts to facilitate an understanding of the process of clinical practice, using critical analysis and interpretation through its incorporation in the teaching of family medicine (Lichtenstein, 2000; Shapiro & Talbot, 1991). By participating in these activities, supervisors are helping trainees to model the skills of self-assessment as part of the process of reflective practice. In practice, reflection-in-action often requires the learner to adjust to the idiosyncratic characteristics of clinical practice. A learner can provide a textbook response, or form a hypothesis, improvising to find a solution to elicit feedback from the patient. These strategies widely inform general practice vocational training and also represent the interpretation of reflection-in-action most relevant to clinical practice. They also provide the pathway to

discovering the individual's professional limitations, which might be addressed through further medical education. This is the interpretation of reflection-in-action most relevant to clinical practice.

Summary of Reflective Practice

Exploration of the work of Schön highlights a series of themes relevant to the process of general practice vocational training, and serves to extend further the perspective of the learner. From the case study of the architect's studio, educationalists have inferred that the work of Schön articulated the need for students to command more than a knowledge base, and to embrace the skills associated with the ability of how to learn. The essence of Schön's endeavour therefore, could be considered to be a theory of meta-cognition. In response, medical boards such as the General Medical Council in the United Kingdom have instituted, in recent curriculum development, a requirement for critical thinking and meta-cognition (Southgate et al., 2001).

Although Schön's initial work began in 1978, the need for the reflective practice and reflection-in-action continues to resonate with medical educators. Schön's contribution continues to remain relevant for any learning situations based on the concept of the reflective practitioner that implies the ability to practice self-assessment. The inclusion of Schön's work as a part of the model on page 94 provides the justification for the use of self-assessment as a means with which individuals might question their own strengths and limitations. The role of the supervisor in general practice vocational training is therefore pivotal in this regard.

Adult Learning

According to Fleming (1993) and Hays and Wellard (1998), Knowles was another influential theorist whose work impacted upon general practice vocational

training in Australia. In contrasting adult learning to pedagogy, Knowles (1984) outlined several principles. These included the need to know, the learner's self-concept, the role of the learner's experience, readiness to learn, orientation and motivation. Not only are these principles evident in the course structure, content and training of medical educators of the program, but they are also articulated in continuing education materials aimed at general practitioners after their specialist training (Royal Australian College of General Practitioners, 2002). Links between these principles and the evolution of general practice vocational training in Australia have emerged in previous qualitative research with supervisors. The principles of adult learning are evident in the use of self-assessment in the educational research conducted in Chapter five because they address the skills of how to learn. This examination of the principles of adult learning from general practice supervisor focus groups demonstrate that the relationship between educational practice and theory is not always as smooth as it might be. Nevertheless, the discussion is worth pursuing because of its relevance to the model on page 94 upon which the self-assessment study is based.

The Need to Know

Knowles articulated the principle that learners, particularly adult learners must have a need to know in order to be motivated enough to expend time and effort on the learning process. He also recommended, that learners needed to participate in planning and directing their learning to ensure that the content remains relevant. Knowles' understanding of adult learning was in contrast to a model of pedagogy in which students learn what the teacher deems to be important (1984).

In keeping with Knowles' principles, the founders of general practice vocational training recognized the importance of the need to know or relevance of the learning experience, and the orientation of the program to real-life experience. As key figures in the Family Medicine Programme, Fabb et al., (1976) contrasted the previous educational approach of undergraduate medical education with its rote

learning, the 'jug to mug approach,' to the new discipline of family medicine (as general practice was known at the time). Considerable enthusiasm accompanied this new approach to education in family medicine which might actually have provided the impetus for the founders to disseminate ideas arising from medical education. One of the key tenets arising from general practice vocational training was that by moving from a system of self-teaching to a sophisticated program relevant to learners' needs, learning outcomes would be enhanced. This represented an admirable though considerably idealistic view of learning.

One assumption of the adult learning movement, in its advocacy of self-directed learning, was that the development of informed, critically aware learners would result in wider social change. Of all assumptions about adult learning, this remains the most questionable. Without reiterating all of the arguments against this proposition, one remains powerful: critically aware learning is not necessarily a socially reformist activity, because self-directed learning often focuses on the individual. Despite this, a three-way relationship exists between Knowles' principles (1984) of the need to know or relevance of a learning experience and the orientation of a program to life experience, and the ideals of social improvement espoused by the adult learning movement. In the context of general practice vocational training, experiential learning models represented a way to preserve the ethos of the family practitioner as healer because medical educators, supervisors and faculty members might model such behaviour to their trainees. In addition, a progressive view of general practice vocational training worked in tandem with the theme of learner participation, the encouragement of informed participants (able to assess their own strengths and weaknesses) and a sense of mission to preserve the progressive aspects of the work of the family practitioner (Candy, 1991).

As general practice training moved from a voluntary system to a mandatory one, for those seeking to practice in the 1990s in Australia, emphasis upon Knowles' principle of the need to know underwent change. It might be that learners participated in determining the direction of their learning to a lesser extent. At the same time, workforce recruitment and distribution became priorities of the funding body, the Commonwealth Department of Health. The view of family medicine as essentially altruistic progressively diminished since the Royal Australian College of General Practitioners began vocational training in 1973. This process related to changes connected to the costing of general practice, the industrial issues of the profession, and the changing perceptions of workforce numbers. Higher rebates for vocationally registered doctors, and the need to gain adequate and varied general practice experience to achieve a pass in the summative examination also became critical factors to the model of training as it evolved. These developments signalled a diminution of voluntarism in training on the part of supervisors and trainees.

By the time the evaluation of vocational training was implemented at the Royal Australian College of General Practitioners in 1995, the previous voluntarism among learners and supervisors, an important characteristic of adult learning, had become less prevalent. Trainees had become more pragmatic about what they chose to study; but the need for them to be able to declare areas of deficiency for learning to be relevant to changing circumstances remained.

Learner's Self-concept

Knowles advocated that adult learners view themselves as autonomous and responsible beings. According to this scheme of things, the autonomous status of the learner is a prerequisite for self-directed learning. For this reason, Knowles placed great emphasis on the need for learners to be active, in counterpoint to

being passive recipients of knowledge. This is consistent with the critique of Fabb et al. (1976) of the previous approach of undergraduate medical education and efforts to change this in general practice vocational training.

In focusing upon the learner's self-concept, Knowles (1984) intended to highlight autonomy as distinct from self-teaching; but this does not prevent the ideology of autonomy becoming a limiting factor in the implementation of lifelong learning. The most contentious point was whether to consider self-direction as an adult characteristic. Although it has been accepted that not everyone is a self-directed learner (Knowles, Holton, & Swanson, 1998), autonomy was well-integrated into the model of training adopted in general practice vocational training in Australia. From time to time, however, supervisors reported a level of frustration with their trainees' inabilities to direct or take responsibility for their own learning. Recent theoretical approaches have suggested educators can contribute to an enhanced state of self-directed learning by motivating learners. This is particularly the case if contextual factors, including self-management and proficiency of the learner and facilitator, are incorporated with cognitive factors such as learners taking responsibility for facilitation, and engaging in reflective thinking and other motivational aspects of education (Garrison, 1997). Commentators on general practice vocational training such as Dr Nowotny have suggested that conditions might not always have occurred as intended, and in some part, this resulted in the decision to incorporate the observation of trainees into the program (Wilde, 1998).

Knowles' principle of the learner's self-concept was manifested by attempts within general practice vocational training to encourage mutual planning by supervisors and trainees. It was also part of an attempt to utilize a curriculum essentially driven by patient cases through the implementation of learning plans.

A difficulty appreciated by many educational reformers (including Knowles), is that not all learners are self-directed in their learning style. For example, Kolb (1984) suggested that general practice trainees are convergent learners, with a tendency to abstract conceptualisation, and experimentation, problem-solving and practical application. These traits are all characteristic of an active learning style. Despite this, general practice supervisors reported mixed views about the abilities of their trainees to be self-directed. Part of the process of encouraging active learners entails that they make a thoughtful appraisal of their skills and that their supervisors observe them and provide feedback. There is some evidence that this did not occur as frequently as it might. Supervisors, on one hand, reported high levels of satisfaction with the content and frequency of feedback they provide to trainees (Goldman, 2000b). Trainees, on the other hand, reported that their supervisors' teaching usually consisted of impromptu teaching, rather than the use of the curriculum (Cooper et al., 2001c). This might suggest that not only were trainees not practising self-assessment, but also, supervisors were not as self-aware as they might have been. These findings relate to the concept of vanishing feedback from Ende (1983) which might be explained by a conflict between the principles of autonomous learning, and the pressure of filling service gaps in general practice (Wearne, 2000).

The lack of training of supervisors in providing feedback to trainees about their performance, and the tendency of some supervisors to use trainees to fill service gaps, might also compromise the learner's self-concept in ways not foreseen by Knowles or the founders of general practice vocational training. Most self-assessment relies on feedback to encourage self-directed learning. Although the model in Figure 1 on page 94 does not measure supervisor feedback, it will be considered as part of the perspectives of educators.

The Role of the Learner's Experience—Emphasis Upon Experiential Techniques and Peer Learning

As previously highlighted, Schön's emphasis on the importance of experiential learning, as articulated in *Educating the Reflective Practitioner*, was an essential part of the development of general practice vocational training. Knowles has suggested a number of points about the role and significance of the learner's experience. Adult learners can see themselves as lacking in recognition if their previous experiences remain unacknowledged. Peer learning, an important component of the learner's experience, according to Knowles, was evident in the education releases of general practice vocational training. Educational releases were sessions in which trainees in regional groupings met with medical educators for educational activities.

The role of the learner's experience remains significant in general practice vocational training due to the type of apprenticeship model adopted. The supervisor carried the responsibility to ensure that the mix of patient presentations provided adequate experience and that they appropriately fulfilled the role of critical friend providing adequate feedback to the trainee. Yet, the question of how prepared supervisors might be for this role remains. They were not educated in how they might foster trainees' critical appraisal skills or indeed their own. The question arises as to whether someone who cannot practise self-assessment will be able to teach others the skill. In contrast to medical educators who have educational experience, supervisors were not provided with educational support until reasonably late in the development of general practice vocational training. Few supervisors, for example, had educational qualifications beyond their own certification from the Fellowship of the Royal Australian College of General Practitioners. Those who had educational qualifications had taken the initiative to seek them. Although later recognition of supervisors extended their role and provided hints to them about teaching and learning (Royal Australian

College of General Practitioners, 1999a), supervisors were never accorded the same status as either trainees or medical educators in the Training Program, and this potentially affected the quality of supervision of trainees. Nevertheless, had general practice vocational training continued in the same way; this aspect of it might have improved.

Readiness to Learn

According to the adult learning model, educators deal with diverse groups of learners, with some requiring more or less support or direction. Ideally, adult learners derive the benefits of learning experiences when they recognise their own readiness to learn. It is also acknowledged that the principle of readiness to learn can be encouraged (Brookfield, 1986; Knowles, 1984). Of most relevance to general practice vocational training is whether general practice supervisors were encouraged, or sufficiently equipped to stimulate readiness to learn in those trainees who were less than responsive. Their lack of experience in knowing that part of readiness to learn entails understanding what one needs to learn, is evident in some preliminary qualitative research on supervisors' views of general practice vocational training (Goldman, 2000a). The focus group produced the following finding. Supervisors reported a series of issues affecting the preparedness of trainees to learn. This was in particular reference to the transition from voluntary to mandatory vocational training for recognition as a general practitioner in Australia. Supervisors viewed trainees as having become more goal-oriented, rather than inspired (or even taking responsibility for their own education). There was no reference to the need to help trainees in gaining the skills of self-assessment. The focus was on conditions external to learning. One supervisor in the focus group responded to the changing circumstances with the following comments:

One of the big things we are focusing on is that our enthusiasm doesn't always match the registrars' enthusiasm. The registrars' enthusiasm is very variable... They are often in the country, and that is not where they want to be. So they are only going to be enthusiastic for what they want... We've all been there, done that, and if there is an exam looming, that can be an issue. And if it isn't relevant to the exam then they're not interested.

I've tailored my enthusiasm to the enthusiasm of the junior registrar. I welcome them initially and ask what do you want to do? They say they want "office-based medicine close to the city"... And then there are the other ones, [who say] "OK, we want to do this and this". I enjoy them much more. We know how to trade off about that. But that's the way it works, and I think we discuss that initially (Goldman, 2000a, p. 5).

In this statement, the supervisor addressed the contrasting enthusiasm of trainees, but also reflected upon the way that workforce issues interact with the system of vocational training. From 1995, the Commonwealth introduced a rural training stream in Australia, requiring that trainees spend a year of their training in a rural area. Educational principles were not necessarily driving this decision. Along with other decisions, political imperatives often prevailed. Although there were efforts to conduct the Training Program along educationally sound principles, and the role of the Education Committee of the Royal Australian College of General Practitioners was significant, trainees were not always appreciative of the requirement to relocate to a rural setting for part of their training.

Orientation to Learning

Knowles contended that adults respond better to learning when presented with life situations. He contrasted the rote learning of rules and principles, in

courses on literature, to the authenticity of students writing about their own experiences. Similarly, the emphasis on authentic life situations with patients as the cornerstone of training addressed the orientation of the learner as a key principle of adult learning, and underpinned the curriculum of Australian general practice vocational training. It represented an ideal opportunity to incorporate self-assessment as the initial step before planning which parts of the curriculum to address.

Motivation

The power of internal motivation, as a stimulus for positive learning experiences, is another principle of the adult learning model. Knowles (1984) has also recognized that negative self-concepts serve to block effective learning, and drew attention to the role of self-confidence for effective learning experiences. This serves to highlight the importance of self-assessment in education. The focus group supervisor quoted below presented as someone who was initially out of touch with a trainee who was seemingly aloof and left him alone. It emerged that the trainee really lacked confidence. Fortunately, the supervisor identified the lack of confidence in the trainee. By presenting himself as fallible and putting some energy into the process of having a trainee, the supervisor found a basis upon which to motivate the trainee. He also detected a connection between negative self-concept, and internal motivation. Responding to a question about trainees' confidence, the supervisor commented:

Unconfident general practitioner registrars [trainees]—they catch you by surprise sometimes...It takes you about a month to work out that they are not really sure what they are doing...The ones who lack confidence, I think that first you have to show them that one, you are not one hundred percent sure of what is going on in your own practice, but also that you care for them and are happy

to spend some time with them to start with. That's what we have done initially with those ones (Goldman, 2000a, p. 5).

Similarly, Garrison (1997) argued that motivation was a key element of learning in medical education, and highlighted the importance of authentic interest and desire to construct personal meaning on the part of learners, and facilitate what he termed worthwhile knowledge. The educator's task, therefore, was a matter of understanding the conditions around motivation, in order to enhance self-directed learning. Part of understanding the concept of motivation for supervisors meant accepting their own strengths and weaknesses.

This discussion of Knowles' principles demonstrates that the relevance and orientation of learning, importance of student autonomy, and role of the learner's experience were all features that general practice vocational training sought to address. These formed the starting point for learning. In the beginning, learning was oriented around the ethos of family medicine. This more altruistic focus gave way to industrially based considerations—issues such as workforce distribution, as well as the introduction of summative assessment and vocational registration. These factors did not necessarily serve to support self-assessment skills in supervisors or trainees and affected general practice vocational training in Australia with regard to the implementation of Knowles' principles; however, they do serve to provide a background for the self-assessment study in Chapter five.

Conclusion on Influential Theories on General Practice Vocational Training

Three inter-related trends in relation to various proponents of adult and experiential learning have been discussed. They served to highlight the origins and values of experiential learning and the need for self-assessment so evident in the work of the founders of general practice vocational training. The need for learner participation and the recognition of what one needs to learn were integral

to the emergence of self-directed learning, particularly in the model of training adopted for general practice training in Australia. It was in this context that objective assessments and subjective assessments of learners (gained through self-assessment) combined to generate an informed understanding of educational processes. All of these factors are fundamental to an understanding of the self-assessment study presented in Chapters five, six and seven.

Also considered, are theories-in-use in general practice vocational training in Australia, and the influential work of both Schön and Knowles. The substantive contribution of experiential learning, arising from its origins in both rationalism and empiricism, was the recognition of the learner's experience and personal values as important (Knowles, 1984; Schön, 1987). In *Educating the Reflective Practitioner*, Schön placed the learner at the centre of the interaction. In terms of general practice vocational training, this recognition of the learner is manifested in the experience of interacting with patients. As such, the curriculum and research literature appropriately acknowledges the importance of patient-general practitioner communication above other areas in the training of general practitioners.

The principle of the learner's self-concept or autonomy has been the key to understanding the work of both Schön and Knowles. In terms of general practice vocational training, the self-direction and autonomy of the learner were important principles in the development of general practice vocational training with regard to reflective practice and reflection-in-action. Several other imperatives, such as workforce and government policy, develop over time and influence general practice vocational training.

This chapter also served to highlight the role of the supervisor in facilitating reflective practice. The mixed reports of supervisors about their trainees, and trainees about their supervisors, suggested that the relationship between

educational theory and educational practice was not always a smooth one. Following Knowles, efforts to encourage supervisors' observations (in order to educate their trainees in self-assessment and provide them with the necessary feedback) were not as effective as they might have been. Similarly, the lack of attention to the role of the supervisor suggested that general practice vocational training could have provided more attention to encouraging supervisors to appraise their own skills.

Later developments within general practice vocational training were in marked contrast to the values of experiential and adult learning as implemented by the founders of the program. These developments sat uncomfortably with the initial basis of the program, where the curriculum "walks in though the door" (Prideaux et al., 1995). In the 1990s, with more prescription about the way that general practice vocational training functioned, there was much less emphasis upon the idealism of the role and importance of education (as a means for changing general practice) than the founders of the program might have wished. The motivation for learning derived from a more singular, external source—the examination. The effect of this shifting emphasis served to make the imperative to learn clear and unambiguous. Despite this change, an emphasis on self-directed learning has remained. Chapter five provides a transition to the self-assessment study at the centre of this dissertation, which depicts the perspective of the learner.

Chapter 5: A Study of Aspects of General Practice Vocational Training

This chapter sets out the aims of a study to evaluate aspects of general practice training. The self-assessment study was to function as an evaluation from the perspectives of learners. It represented the measurable aspects of a model developed from the literature on self-assessment and perspectives of learners and educators as presented in Chapters three and four. This chapter also provides a rationale for the research design, in terms of its relevance to the research questions, and describes the key hypotheses, methods and research procedures followed for the research. Some of the distinguishing features of the research strategy in relation to program evaluation by the Outcomes Evaluation Unit of the Royal Australian College of General Practitioners are also described.

Aims

The main aim of the study was to evaluate aspects of general practice vocational training from the perspective of the learner. Specifically, the study measures, in a longitudinal fashion, the extent to which trainees assessed their own competence across five domains of general practice and the care of particular patient groups and presentations, and trainees' reported satisfaction with aspects of training. The study also measures longitudinal trainee self-assessed competence and satisfaction among a cohort of the same trainees as they progressed through the program.

Finally, the study examines the relationship between self-assessed competence and satisfaction using a model developed from the work of a significant medical educator, as well as other self-assessment literature, educational theory and

qualitative research on supervisors and medical educators. The aims of the study related most closely to the measurable aspects of the model are outlined in Figure 1, page 94.

Design Considerations

Researchers seeking to examine the effectiveness and value of training face a broad range of possibilities (from both methodological and practical standpoints) in their choice of research design. A starting point for a discussion of methodology lies with the research questions and their context. When devising the research questions and designing the research methodology, in order to evaluate even one or two aspects of a program of the scope of the Australian general practice vocational training program, practical concerns were ever present.

Recently, organizations delivering Australian government-funded programs were encouraged to address the effectiveness of outcomes in their evaluation of programs. In this context, outcomes evaluation expanded upon process-oriented evaluation in the belief that the use of this terminology and its shift in emphasis would also affect program outcomes.

A qualitative study might have been a valuable contribution towards an examination of the in-depth, subjective experiences of trainees. Notwithstanding the appeal of such an approach (to a person with a predominantly qualitative research background) there were several reasons not to pursue this course of action. Such a study would not necessarily have satisfied the requirement to focus on research questions around the effectiveness and outcomes of the program. This was a clear need of both the Commonwealth Department of Health and Aged Care and the Royal Australian College of General Practitioners.

Due to logistical factors, a longitudinal qualitative approach is difficult to implement. Unlike a longitudinal study, a one-shot qualitative study would have produced findings limited only to one point in time. This represented a potential difficulty for any program seeking to implement changes arising from an evaluation, as the robustness of the findings could be open to challenge. A further limitation of the qualitative approach was the absence of statistical inference. Medical stakeholders, in particular, might wonder whether it was reasonable to infer from the findings of a qualitative study to a broader population of trainees. As an alternative methodological design option, a randomised controlled trial would be desirable to answer questions as to the extent of the appropriateness of the program. Unfortunately, this form of design was not feasible because of the lack of a control group. Feinstein (1996) provided a salutary warning (that impacted upon this decision) when he recommended that a research design should be feasible, and that researchers should not be overzealous in their choice of research design. This advice was particularly salient to the Australian situation. Since vocational training has become the major pathway to general practice, it has become impossible to enlist a control group of non-trainee doctors with similar characteristics to general practice trainees.

The alternative approach to a qualitative one was a longitudinal study treating the curriculum as the most substantial part of the intervention to be tested. The general practice vocational program was a notionally three-year-program that featured multiple elements across a range of clinical settings. Although there were time allocations for learning activities, the program did not function as a strictly time-based program in which learners participate in a linear sequence of subjects or educational activities. Nevertheless, there were reasons to treat the program as an intervention. The program was coordinated nationally, according to set requirements for teaching and supervision, and delivered at a state or regional level. The structure of the program featured a learning sequence where

the different stages of the program could justifiably be regarded as cohorts. The existence of these cohorts facilitates social research. Most important, the curriculum was considered to be the major part of the program (Kiley, Peterson, Mullins, & Wellard, 2000). This lent itself to the development of educational research. For these reasons, relating to both program structure and content, it was both defensible and reasonable to treat the program as an intervention. As Norman (2002) highlighted, much progress had been made in terms of basic educational research in relation to methodologies, the learning process, and assessment practices. Against the background of these developments, it was reasonable to draw on this progress to evaluate a program of the scope of general practice vocational training in Australia.

In the case of general practice vocational training evaluation, a longitudinal survey approach was a compromise between the naturalistic setting of the evaluation and the more rigorous, but less feasible, randomised controlled trial. Arnold et al. (1985), Donnelly et al. (1994), Gordon (1991; 1992) and Speechley (1993; 1994) trialled the testing of a self-assessment model, and their approach offered a methodologically defensible method that was consistent with the utilization-focused evaluation of Patton (1986).

The relationship between program outcomes and the self-effectiveness of the learner was particularly relevant, given the history of general practice training (as described in the evaluation approach in the Introduction and in Chapter two). Although this relationship suggested an array of possibilities for devising questions for the learner about the value of experiential learning, it would have been difficult to gain support for this as a project. This was because conducting large numbers of focus groups, or arranging interviews in a country as geographically dispersed as Australia would disrupt the education of trainees. On the other hand, it was feasible to conduct a survey with potentially good response

rates for trainees spread across Australia, because a simple cover letter could inform trainees that participation in the survey would contribute to an enhanced training program, which, in turn, would be of benefit for future trainees. The questionnaire method utilized in the study enabled the measurement of a number of variables that could not only be treated as the outcomes of general practice vocational training, but could also generate data that might be generalized to the population of general practice vocational trainees. The possibility of linking program outcomes and trainees self-assessments would provide answerable questions for the trainees, and offer significant critique, feedback and suggestions for program developers, medical educators and supervisors. These possibilities provided the main rationale for the particular longitudinal research design developed and applied to the following study.

Research Strategy

Chapter three examined the perspectives of learners from previous empirical studies. Chapter four examined the perspectives of medical educators and supervisors in relation to educational theories. This chapter explains some of the associated methodological issues of the self-assessment study. One of these aspects relates to the specific characteristics and relationship that developed between the self-assessment study at the centre of this dissertation, and the overall program evaluation of the Outcomes Evaluation Unit. The self-assessment study sought to measure and consider the attitudes of learners to their program. It not only contributed to a portfolio of projects conducted by the Outcomes Evaluation Unit to depict a composite picture of the appropriateness of general practice vocational training; but also functioned quite separately from the program evaluation. The self-assessment study arose from a model that expanded on the literature around the relationship between self-assessed competence and

satisfaction. The program evaluation work within the Outcomes Evaluation Unit did not encompass that model, and it is in this area that the differences between the program evaluation and the self-assessment study were clearest.

In the initial design of the research questions in the self-assessment study, I decided to use questionnaire data as trainees progressed through training. This allowed data to be collected during the program evaluation by the Outcomes Evaluation Unit. This was a pragmatic decision motivated by the possibility of undertaking a project of wider scope than would have been possible under normal circumstances. The consequence of this was that the primary data collected for the study originated from self-assessed survey questions from general practice trainees. (Supervisors also provided a small amount of primary data, and Chapter four refers to this.)

To augment the primary self-assessment data, I drew upon a rich collection of findings from a diverse group of projects conducted during the life of the Outcomes Evaluation Unit. The problem of educational evaluation represented a challenge because of the multiplicity of disciplines it encompasses. In the Five-Year Strategic Plan, developed by the Outcomes Evaluation Unit, there were over 20 projects, designed to measure the broad topics of effectiveness, efficiency and appropriateness of general practice training. The approach nominated in the Outcomes Evaluation Unit contributed to projects on selection processes, practice patterns against standards, self-assessed competence, satisfaction of trainees and their supervisors, and career progression. These diverse projects augment one another in a similar approach to the approach developed by Tamblyn (1994), or that of Denzin (1989) drawing on multiple data sources to corroborate the findings of multiple projects. Although educators favour approaches utilizing multiple perspectives in which different data sources might be triangulated (Prideaux, 2002), such designs are rare and are therefore considered as valuable

and ground-breaking. Perhaps, inevitably, part of the boundary-breaking aspect of such designs is that it is difficult to bring together disparate data sources into a coherent picture. Remaining specific in the hypotheses and realistic about the contribution of the project that might be expected helped to overcome this difficulty.

Another possible approach to program appropriateness might have been to undertake a meta-analysis; but considerable scepticism existed (and continues) as to whether a meta-analysis in educational evaluation is actually feasible. How to assess program effectiveness remains a matter of interpretation, as meta-analyses differ according to the discipline one is working in. In social science, for example, meta-analysis is a circumscribed statistical technique, whereas in education, meta-analysis has a looser definition. This difference in methodologies and their interpretations is sufficient to make the technique indecipherable to an evaluation audience, and this becomes potentially problematic.

Although educational interventions in real settings (as described in Chapter three) are not simple to evaluate because of the time delay between intervention and outcome, Norman (2002) has suggested that along with changes to assessment practices, considerable progress has been made with educational research. These studies are drawn upon and continued in this research by accepting the premise that assessing program appropriateness can be undertaken by examining the available research data and evaluating aspects of the program; namely its ability to foster lifelong learning and reflective practice in learners. The self-assessment study at the core of this dissertation represents the empirical research data and analyses derived from the perspective of general practice trainees, and will be compared and put in the context of the perspectives of educational theorists.

Hypotheses

Whereas Chapter three emphasized the perspectives of the learner, Chapter four provided an insight into the educational context for the perspectives of medical educators and supervisors on student self-assessment. The combination of these factors provided the background and rationale for developing a self-assessment study to examine aspects of the value and appropriateness of general practice vocational training. An important aspect of this rationale is that self-directed learning has been highlighted by Prideaux et al. (1995) as an implied outcome of training. The findings of several studies of self-assessed competence and satisfaction provided guidance on some relevant and testable hypotheses (Goldman, Jasper, & Wellard, 1998). The findings on gender and locality were supported by the findings of Goldman et al. (2001a). Arnold's study (1985) and a later meta-analysis of Falchikov and Boud (1989) found a significant improvement in self-assessed competence as trainees progressed. These studies support the stage of training hypothesis.

The hypotheses examined in the self-assessment study arising from the research questions in the description of the self-assessment study in the Introduction were:

Hypothesis 1: Self-assessed competence ratings in curriculum items can be anticipated to increase as trainees progress through training (from Chapter three).

Hypothesis 2: Male and female trainees can be predicted to report similar attitudes with regard to self-assessed competence in curriculum items (from the Introduction and Chapter three).

Hypothesis 3: Urban and rural trainees can be predicted to report similar attitudes in self-assessed competence (from Chapter three).

Hypothesis 4: Self-assessed competence ratings can be expected to decrease once trainees see a certain number of patients (from the Introduction).

Hypothesis 5: Self-assessed competence of a longitudinal cohort of trainees can be expected to increase as they progress through the program (from the Introduction and Chapter three).

Hypothesis 6: Satisfaction can be anticipated to increase as trainees progress through training (from the Introduction and Chapter three).

Hypothesis 7: Male and female trainees can be predicted to report similar attitudes with regard to satisfaction with training (from the Introduction and Chapter three).

Hypothesis 8: Urban and rural trainees can also be predicted to report similar attitudes with regard to satisfaction (from the Introduction and Chapter three).

Hypothesis 9: Satisfaction ratings can be expected to decrease once trainees see a certain number of patients (from the Introduction and Chapter three).

Hypothesis 10: Satisfaction among a longitudinal cohort of trainees can be expected to increase as they progress through the program (from Chapter four).

Hypothesis 11: The extent to which trainees are satisfied with the program of training might be expected to determine their assessments of competence. The hypothesis to be tested was generated from an analysis of the literature on the use of self-assessment in Chapter three and previous research from the author (Goldman, 1999).

Hypothesis 12: Trainees' assessments of their own competence will be likely to display little concordance with examination results (from Chapter three).

Hypothesis 13: Self-assessment will be a useful tool in relation to general practice training when considered in relation to educational theory on reflective practice (from Chapter four).

Ethics

The Ethics Committees of the Royal Australian College of General Practitioners and Deakin University granted approval for the conduct of this study. Upon transfer of candidature from Deakin University, the University of Tasmania Ethics Committee provided ethics approval for the study (approval number H000631) on 12 November 2001. Research assistants did not report any issues of importance to the ethical conduct of research when conducting telephone follow-up of non-respondents. Confidentiality of questionnaire data was maintained.

Method

Using NVivo qualitative research software (Bazeley & Richards, 2000) an analysis was conducted on studies examining the appropriateness of training from the perspective of learners. Primarily, these studies asked trainees to identify their strengths and weaknesses, and preparedness to practice (Biro et al., 1993; Calhoun et al., 1990; Clack & Head, 1999; Cochran & Spears, 1980; Fincher et al., 1993; Keck & Arnold, 1979; Kolm & Verhulst, 1987; Morton & Macbeth, 1977; Sclabassi & Woelfel, 1984; Speechley et al., 1993). The self-assessment study at the centre of this dissertation was informed by these quantitative studies, and extended a particular longitudinal cohort study from Arnold et al. (1985). This study was the starting point for the development of a model. The model extends this work by depicting multiple perspectives of self-assessment in general practice vocational training. The model comprises both measurable and observable characteristics in the relationship between self-assessed competence and satisfaction as depicted in Figure 1 on page 94. These features defined the study as an educational evaluation. The model also has a longitudinal aspect in as far as the self-assessment study has a cohort of participants in the study over three years.

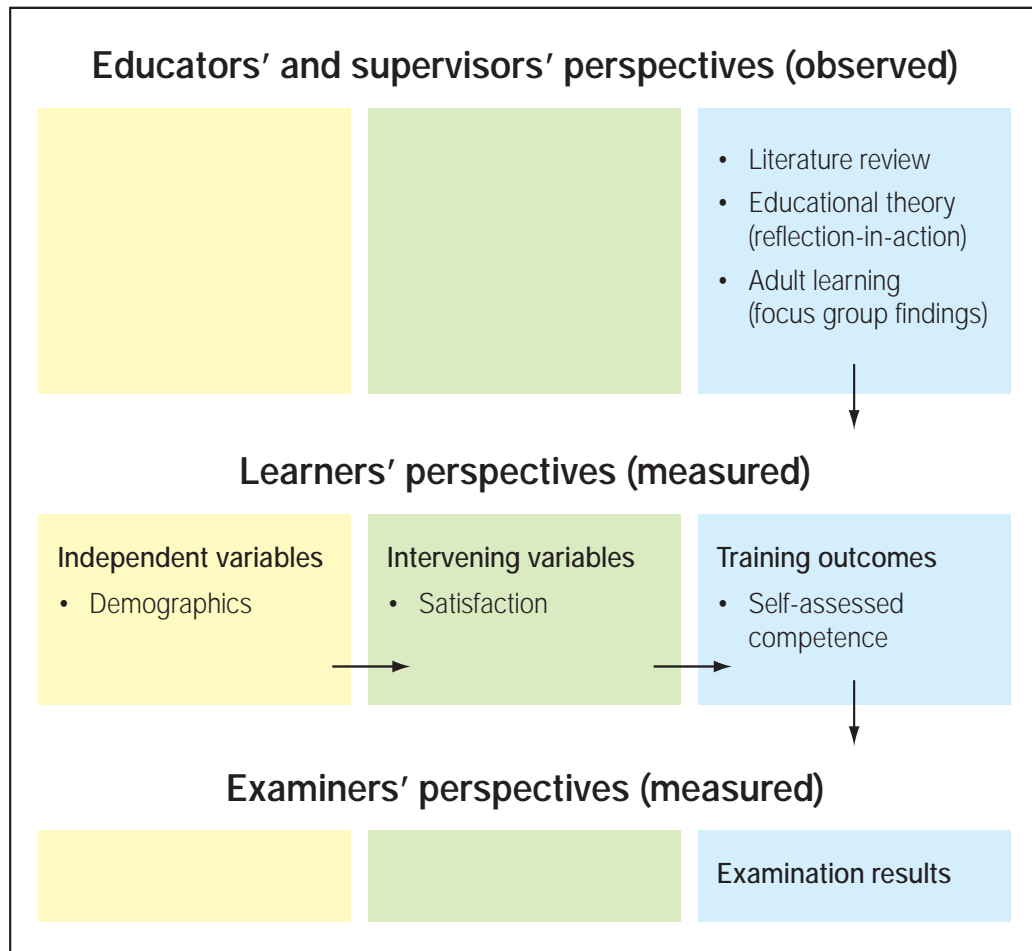


Figure 1: Measured and observed aspects of the relationship between self-assessed competence and satisfaction

In Chapter four, a discussion of relevant trends in professional self-examination occurred in conjunction with qualitative research on medical educators' and supervisors' views. This discussion provided the context and rationale for the educators' perspective as depicted in the educational theory and adult learning aspects of the model. Observations around professional self-examination and life-long learning also feature here. These relate to the observable aspects of the model. The other perspectives on self-assessment derive from those of learners. The learners' perspectives are represented in the measured aspects of the model in relation to the self-assessment study in Chapter five. Additional variables

that explained patterns observed in other studies consist of the independent demographic variables such as the stage of training, gender, and locality. As has already been touched upon, the stage of training variable referred to the different stages over the lifecycle of a general practice trainee, whether hospital training over one year, the basic, advanced general practice terms, subsequent general practice experience or two years after the completion of the program. This variable facilitated a key part of this study to evaluate the appropriateness of the program from trainees' perspectives by testing the hypothesis that trainee self-assessed ratings would increase as trainees progressed.

The variable of gender was included as a demographic variable because the literature has placed some emphasis upon it. Britt et al. (2000) reported that female general practitioners had different practice populations from males, and therefore were presented with different clinical problems. Clack and Head (1999) and Roter, Hall and Aoki (2002) found that females managed their patients in different ways from males and exhibited different personal attributes. Johnson et al. (1993) drew attention to the different career patterns of women and men. With an increasing proportion of female general practice trainees, a more sophisticated understanding of the effects of gender is necessary. The changing gender balance has potential implications for the general practice workforce of the future.

Locality was an important feature of policy making for general practice vocational training throughout the 1990s in Australia because of its relation to workforce planning. The Commonwealth Department of Health and Family Services made many attempts to address the shortage of rural practitioners in general practice vocational training through interventions such as the rural training stream. A report written on both locality and gender using the same data as Chapters six and seven produced some unanticipated, significant findings (Goldman et al., 2001a). The report found that a trainee's gender was a factor that

influenced self-assessments, whether trainees were from urban or rural areas. This suggests that initiatives targeting the issue of locality also need to consider gender in their planning.

The number of patients seen by trainees was also a demographic variable in the model. The importance of workload, in providing a balance between the number of patients and educational activity was based on the work of Fleming (1986). General practice vocational training had established particular requirements in order to facilitate the optimal balance between clinical and educational releases and to ensure that supervisors, as employers of trainees did not exploit them. It also stipulated that students needed to see a sufficient number of patients to enhance their skills.

Satisfaction was the intervening variable in the model and this included items on communication factors and work conditions. Satisfaction with communication factors included opportunities to learn about communication and public health, engage in clinical teaching sessions, and participate in video consultation reviews. Satisfaction with work conditions included satisfaction with location of training posts, remuneration and allowances, other terms and conditions, and the level of workload expected. Satisfaction was particularly important in initiatives that supported self-directed learning—a central aspect of general practice vocational training. Bradley et al. (1996) set a precedent in this regard with the use of satisfaction data to develop recommendations about changes to general practice vocational training in Britain. Although students might not necessarily be the best judges of their overall educational requirements, data on satisfaction with general practice, provide an understanding of how to enhance the curriculum, and the overall training experience.

The dependent variables in the model of trainees' self-assessed competence related to areas of the vocational training curriculum. These included the five domains of general practice, and particular patient groups and presentations. As demonstrated in Chapter three, Donnelly et al. (1994) found self-assessment to be a reliable and valid outcome measure for the evaluation of training. Other studies of general practice education programs also used the self-assessment of graduates as a means to gain insights into their preparedness to practice (Arnold et al., 1985; Goedhuys & Leroy, 1990; Hojat et al., 1996; Hojat, Vogel, Zeleznik, & Borenstein, 1988; Sciabassi & Woelfel, 1984). The work of Bligh and Slade (1996) was noteworthy for highlighting the importance of learner-centredness in medical education, through the use of a self-directed learning readiness scale in six regions of the United Kingdom. These studies provided examples using similar models to the study outlined here.

The last aspect of the model relates to the perspectives of examiners represented by examination results. These are compared to self-assessments to examine potential correlations.

In summary, the model depicts the relationship between self-assessed competence and satisfaction using demographic variables such as stage of training, gender, locality, patient population and workload, as well as satisfaction with aspects of the program and self-assessment in curriculum items based on a summary of the literature review in Appendix A 1. These were the measurable variables. The intervening variables included examination results, as a measurable variable, and clinical feedback from supervisors, as an observed variable. Supervisor data from Chapter four on reflective practice augmented the conceptual basis of the model. Self-assessed competence and satisfaction with various aspects of training were the outcome variables. Their purpose was to identify aspects of training that might explain the appropriateness of training for the learner.

Procedure

With the advent of the Evaluation Directorate, robust dialogue began about which evaluation questions might demonstrate the effectiveness of general practice vocational training. Wilde (1998) in her history of general practice vocational training, suggested that a tension was evident between political and educational representatives within the Royal Australian College of General Practitioners. John Docker was the Head of the Evaluation Directorate of the Royal Australian College of General Practitioners. The approach of the Evaluation Directorate was exemplified in a report on knowledge, attitudes and skills as a precursor to practice behaviour (Docker et al., 1993). Although the publications from the Evaluation Directorate were numerous, difficult relationships developed between the Evaluation Directorate and the Training Program management of the College. Ian Steven was the subsequent Head of the Evaluation Directorate, and his work on knowledge testing and audit (1991) represented a stark contrast to the philosophy of education as a means to change knowledge, attitudes and skills. Despite his considerable influence on general practice research, a coherent evaluation plan for general practice vocational training was neither developed nor implemented.

The previous less than successful efforts to start an evaluation program of general practice training served to slightly constrain evaluation activities. After these various attempts to initiate the evaluation of general practice vocational training, Rod Wellard (at that time Director of Education), accepted the role of evaluation manager. At that time, evaluation began to function independently of, but collaboratively with, the Royal Australian College of General Practitioners and the Training Program to make for a workable arrangement.

A comprehensive program of evaluation was proposed to measure the effectiveness, efficiency and appropriateness (or value) of training, based on the

concept of utilization-focused evaluation from Patton (1991). The evaluation plan, endorsed by the Outcomes Evaluation Advisory Committee that had oversight of the Outcomes Evaluation Unit, articulated this approach. The program of evaluation was to have clear outcomes rather than being open-ended, as might have been the case with a more educational focus.

Educators of the Training Program also recognized the legacy of the program founders, in so far as without a sense of preparedness to practice or a readiness for self-directed learning, the program would not be meeting its objectives. As a response, one of the projects in the evaluation plan was a longitudinal cohort study on the self-assessed competence or preparedness of trainees for general practice. The self-assessment study sought to address one of the earliest objectives of training, which was to foster self-directed learning; however, it should be emphasized that trainees did not receive any specific training in self-assessment techniques.

To test the relationship between self-assessed competence and satisfaction, this study comprised a longitudinal panel study with annual surveys from 1999 to 2001. The two main concepts contained within the questionnaire were self-assessed competence or preparedness to practice, and satisfaction with aspects of training. These concepts were developed based on the findings of a pilot study conducted during the previous year (Goldman et al., 1998).

The first annual survey was conducted in March 1999. The second annual survey of August 2000 and the third annual survey of August 2001 were conducted with the previous sample, in addition to some additional recruits. A longitudinal cohort comprised a group of participants who had completed each of the three surveys. This analysis was undertaken retrospectively. For year two, 161 trainees had completed two surveys, and by year three 134 trainees had completed each of the three surveys. Cross-sectional and longitudinal analyses were used to examine

trends over time, and to ascertain whether the cross-sectional results among the sample of the survey respondents remained consistent.

A novel aspect of the model was in its longitudinal capacity to follow a cohort of individuals across three surveys. This was to provide insights otherwise unavailable through cross-sectional studies. This longitudinal approach served to make the findings on self-assessment both robust and amenable to comparison with similar studies of trends over time. Notwithstanding the exploratory role of self-assessment, the logic underpinning this type of research design was based upon the assumption that although respondents might both under- and over-report in their self-assessments, they would do so with reasonable consistency. Similar logic underpinned the collection of self-reported health data in the studies of Glasgow et al. (2005) and Siahpush, Borland and Scollo (2002).

Questionnaire Development

A search of relevant literature confirmed a lack of standardized items for collecting data on a general practice vocational training program—especially one with the scope of the Training Program. This finding was consistent with Prideaux et al. (1999) who observed that home grown questionnaires in the area were plentiful. This feature might have been due to the specialized nature of vocational training curricula. Arising from the use of such questionnaires was the issue of validity. As important guides through the vast array of questionnaires that could be used in the area of clinical practice, Feinstein (1996) and Allen and Rashid (1996) have usefully suggested the need to focus on face validity, content validity and construct validity.

In the context of general practice vocational training, the process used to develop the questionnaire addressed the matter of content validity. By aligning the questionnaire to the curriculum framework, the questionnaire was based on

previous research on the prevalence of patient presentations in Australian general practice (Bridges-Webb, Britt, Miles et al., 1993). This process ensured that the five domains of general practice and the sampling of clinical problems addressed in the questionnaire broadly reflected everyday practice.

The face validity of the questionnaire referred to its ability to address areas of considerable relevance to general practice. Bearing in mind the scope of this study, some pragmatism was required in this regard, for the construction of general practice competencies might have been a whole project in itself. Two elements suggested the face validity of the questionnaire. First, the questionnaire mirrored the general practice vocational training curriculum, and was therefore a legitimate item for testing. This was due to the process of development of the curriculum, involving an extensive set of national consultations to gain the substantial input of medical educators and general practitioners (Kiley et al., 2000). Second, the process for review of the questionnaire was inclusive. It featured the input of representatives of the Royal Australian College of General Practitioners, the General Practice Registrar Association and the Outcomes Evaluation Advisory Committee, and general practitioners and supervisors of trainees. Their review of the questionnaire provided a check to ensure that the questionnaire represented key areas of general practice. In addition, findings from two general practitioner focus groups helped to ensure that the questionnaire was valid from the perspective of those working in the field of everyday general practice.

Construct validity was defined by de Vaus (2002) as referring to the issue of whether questionnaire items confirmed theoretical expectations. The work of Arnold, Willoughby and Calkins (1984) was drawn on to examine expectations of similar data. They found an association between trainees' stage of training and self-assessed competence. Similarly, the pilot study to examine general practice

vocational training in Australia, found that trainees in later stages of the program reported higher self-assessed competence scores than those at earlier stages (Goldman et al., 1998).

By recognizing and addressing these issues in the development of the questionnaire, it was therefore possible to claim that it possessed content, face and construct validity.

As can be seen in Appendix D 1, most of the items in the questionnaire were in the form of Likert scales. Questions concerning competence asked respondents to agree or disagree that they were competent in a range of activities. This was to make it easier for respondents to concede areas of weakness, rather than directly asking them whether they were competent or not. The scales ranged from 1–5 (1=Strongly disagree and 5=Strongly agree) with a neutral mid-point. Questions concerning satisfaction have a 5-point scale ranging from (1=Very dissatisfied and 5=Very satisfied) with a neutral mid-point. Competence and satisfaction questions were converted to a scale out of a hundred to assist with readability. Longitudinal variables were collapsed into three categories (1=Agree, 2=Neutral, 3=Disagree) for each of the surveys.

Decisions about placing questionnaire items into the appropriate domains of general practice involved some judgment. This was assisted by comparing the five domains of general practice with the work of Hays, van der Vleuten, Fabb and Spike (1995) on the reliability of the Fellowship Examination of the Royal Australian College of General Practitioners. In addition, the process of curriculum development was a dynamic one. For this reason, there were some minor changes, which occurred over time, to the five domains of general practice. (Appendix C 1 shows the questionnaire items as they related to the five domains of general practice.)

Sample

The enrolment database of the Royal Australian College of General Practitioners provided a list of the total population of trainees in the hospital, basic and advanced stages of the program, and trainees engaged in subsequent general practice experience. Another group of trainees who had previously completed their training was also included.

For year one, the total possible population was 1565 current and former trainees who had previously completed their training in 1997. The sampling software *Epi Info* (Dean, 1996) indicated that 770 respondents would be sufficient for the required statistical processes to measure an effect size of around 10 percent in self-assessed competence and satisfaction variables. By the time of the second survey, limitations by the Department of Health and Family Services on the numbers of general practice trainees took effect, and the total possible population of trainees was 1366. Statistical advice suggested that for this survey, 550 current and former trainees were sufficient. Former trainees of that year had completed their training in 1998. During year three, the total possible population of trainees was 1062. Of this population, advice suggested that 450 trainees were sufficient for the study. In that year, former trainees had completed their training in 1999.

The method used to draw a sample from the population of trainees was as recommended by de Vaus (2002) for drawing stratified random samples. This consisted of drawing a list of randomly selected cases each year, and examining the variables of stage of training and state to ensure that the sample was representative of the total trainee population.

Recruitment

To meet the inclusion criteria of the study, respondents were required to be in the hospital year, the basic or advanced term, or doing subsequent general

practice experience. Also eligible was another cohort of doctors within two years of having completed their training. Only doctors on leave were excluded. During each survey, respondents received a cover letter informing them about the purpose of the project, its voluntary nature, and issues including the confidentiality of data. Initial non-respondents received up to four telephone calls to encourage them to return questionnaires.

Questionnaire Data Analyses

Data were analyzed using SPSS (Norusis, 2002). Analyses of variance (ANOVA) or *t* tests were used and mean scores compared to test the first and second hypotheses: that self-assessed competence and satisfaction with aspects of the curriculum increase as trainees progress through training. The choice to use mean scores was based upon the large sample sizes in the study and the need to provide ready comparisons of education items. (For years one to three of the survey, *n*=621, 549 and 472 respectively.) In conjunction with these tests, post hoc tests were used to decide which results to highlight. The decision to use Scheffé test as a test of significance for post hoc tests in preference to the Bonferroni test was based on the appropriateness for Scheffé for the social and psychological sciences. It was also thought that the precision of the Bonferroni test might lead to Type II error (or a situation in which it is falsely accepted that there are no differences between groups). Bonferroni tests for the domains of general practice items produced the same results as Scheffé post hoc tests. The initial decision was therefore appropriate. Scheffé tests were also used to examine the third and fourth hypotheses on gender differences in self-assessed competence and satisfaction, the fifth and sixth hypotheses on locality differences, and the seventh hypothesis on workload differences in self-assessed competence and satisfaction. Repeated measures ANOVA tests were used to analyze the findings of a small cohort of trainees completing each survey, to determine whether trends identified among the cross-sectional annual survey results were consistent.

The longitudinal model in Figure 1 on page 94 was expected to provide an understanding of the extent to which trainees' satisfaction might explain the extent to which they viewed themselves as competent. For the sake of keeping reporting succinct, scales were used. In order to describe trainee satisfaction in specific terms, an additive scale was developed for use in a regression analysis.

For year one, four scales were created, and reliability calculated to ascertain whether the scales measured the same underlying construct. The first scale, work conditions, consisted of satisfaction with location of training posts, remuneration and allowances received, level of workload, other terms and conditions and administration. It had high reliability (Cronbach's alpha .80). The second scale, program components, consisted of satisfaction with access to educational materials, opportunities to meet other trainees, external clinical teaching sessions, video and audio consultation reviews, the logbook, and information about the program. This scale also had a reasonably high Cronbach's alpha of .72. The third scale, teaching outcomes, consisted of satisfaction with the role models provided by supervisors, supervision and advice received and opportunities to learn communication skills. This scale, too, had high reliability (Cronbach's alpha .80). The fourth scale, learning outcomes, consisted of satisfaction with the range of clinical experiences and learning challenges provided and had reasonably high reliability (Cronbach's alpha .73).

The second year of the questionnaire produced a different set of scales for use within the regression analysis. In this case, four dimensions of satisfaction were used: satisfaction with work conditions, teaching challenges, communication, and administration. The first scale for this year represented work conditions and consisted of satisfaction with location of training posts, remuneration and allowances received, other terms and conditions, and the level of workload expected. The first scale had high reliability (Cronbach's alpha .79). The second

scale, teaching outcomes, included the range of clinical experiences, learning challenges provided, role models provided by supervisors, and supervision and advice received. The second scale had high reliability (Cronbach's alpha .81). The third scale, communication, included opportunities to learn communication skills, meet with other trainees, and engage in external clinical teaching sessions, and video and consultation reviews. It had reasonable reliability (Cronbach's alpha .73). Finally, the fourth scale, administrative items, included access to educational materials and resources, usefulness of the logbook, information, and administration. It had reasonable reliability (Cronbach's alpha .76).

For the third year of the survey, scale one, work conditions, included satisfaction with the location of training posts, remuneration and allowances received, other terms and conditions and the level of workload expected. The scale has good reliability (Cronbach's alpha .70). Scale two, teaching outcomes, consists of satisfaction with clinical experiences, and learning challenges and role models provided by supervisors, and supervision and advice received (Cronbach's alpha .78). It also had good reliability. Scale three, communication, consisted of communication skills, opportunities to meet other trainees, external clinical teaching sessions and video and consultation reviews (Cronbach's alpha .70). It has reasonable reliability. Scale four, administration, consisted of satisfaction with access to educational materials, the logbook, information and administration, and has useable reliability (Cronbach's alpha .67).

Block recursive regression showed the total causal effects of independent variables upon self-assessed competence—the dependent variable (Graetz & McAllister, 1988). The block recursive regression analysis featured the trainee gender, age, full or part-time status, workload, university of graduation, locality and state as independent variables, and trainee satisfaction with conditions of work, communication and administration with stage of training as intervening

variables to explain the extent to which satisfaction explained self-assessed competence. This simple multiple regression was designed to test hypothesis nine, on the extent to which trainee satisfaction could be expected to determine self-assessed competence. Finally, correlations were used to examine the hypothesis 10, on the relationship between self-assessed competence and examination results.

Comparison of Questionnaire and Examination Data

The Censor-in-chief of the Royal Australian College of General Practitioners approved access to examination data for the purpose of a comparison of self-assessment and examination data. The Examination Department of the Royal Australian College of General Practitioners conducted the data match and de-identified the file.

The de-identified file provided variables on the date of examination completion, subject codes, and the average examination result. Some manual selection was required to ensure that the survey data were as close to the examination date as possible. An analysis compared examination data from trainees with self-assessments in the main curriculum areas. Only the main items were comparable between the Fellowship of the Royal Australian College of General Practitioners Examination and the curriculum.

Summary of the Design of the Self-assessment Study

In this description of a research project for studying general practice training, a model on self-assessed competence and satisfaction featured, along with the independent variables of stage of training, gender, locality and workload. The research methodology in relation to relevant debates surrounding it also played a part. A description of the major concepts of the self-assessment study and the previous research on self-assessment as a valid outcome measure of capability

focusing upon learner-centredness featured. Finally, a description on the design and development of the procedures used in the study provided the setting for the next chapter in which the results emerge.

Chapter 6: Study Findings

This chapter examines the findings of the self-assessment study conducted over three years with trainees of the Royal Australian College of General Practitioners. The main results on self-assessed competence and satisfaction are examined in relation to stage of training, gender, locality and workload. The same areas are examined in relation to longitudinal results including a cohort of the same trainees surveyed over three years. Also examined is the relationship between satisfaction and self-assessed competence. In addition, self-assessments are compared with examination data. The findings are examined in relation to an evaluation of general practice vocational training from the perspective of learners.

Characteristics of Respondents

The response rate for each of the three surveys exceeded 80 percent. For years one to three, the numbers of respondents were 621, 549 and 472 respectively. The major demographic characteristics of the respondents for each survey appear in Tables 1 and 2.

The surveys conducted in years one and two produced comparable demographic results to those of the survey conducted in year three. Generally, the sample of respondents across all years mirrored the population of trainees.

Self-assessed Competence

Questions in this area asked trainees about the extent to which they agreed or disagreed that they were competent to engage in general practice activities in an unsupervised manner. The activities covered the five domains of general practice, the management of particular patient groups and presenting problems—all items addressed by the general practice vocational training curriculum and the Fellowship Examination of the Royal Australian College of General Practitioners.

Table 1: Characteristics of current and former trainees

	1999 percent	(N)	2000 percent	(N)	2001 percent	(N)
Stage of training						
Hospital	19.0	(118)	16.6	(91)	11.4	(54)
Basic	18.5	(115)	19.1	(105)	7.8	(37)
Advanced	15.0	(93)	11.8	(65)	11.7	(55)
Subsequent	20.2	(126)	25.0	(137)	27.8	(131)
Former	27.3	(170)	27.5	(151)	41.3	(195)
Gender						
Male	39.2	(241)	41.0	(224)	42.5	(200)
Female	60.8	(374)	59.0	(322)	57.5	(271)
Age						
Under 30 years	44.9	(279)	40.5	(223)	31.0	(146)
30 years and over	55.1	(343)	59.5	(327)	69.0	(325)
Workload						
0–40 patients	12.1	(69)	8.2	(45)	12.6	(56)
41–80 patients	27.0	(154)	24.1	(132)	25.2	(112)
81–120 patients	38.4	(219)	35.9	(197)	36.0	(160)
121–160 patients	16.5	(94)	13.9	(76)	15.7	(70)
161 patients and above	6.1	(35)	17.9	(98)	10.6	(47)

Table 2: Further characteristics of current and former trainees

	1999 percent	(N)	2000 percent	(N)	2001 percent	(N)
Location						
Urban	68.2	(395)	63.8	(254)	66.6	(301)
Rural/remote	31.8	(184)	36.2	(144)	33.4	(151)
Rural training stream						
Yes	19.9	(121)	22.5	(121)	29.7	(138)
No	80.1	(488)	77.5	(417)	70.3	(327)
State or territory						
NSW	33.2	(209)	32.8	(177)	34.0	(160)
Vic	24.4	(150)	23.9	(129)	22.9	(108)
Qld	16.8	(104)	16.5	(89)	17.8	(84)
WA	10.7	(66)	10.0	(54)	9.8	(46)
SA/NT	8.6	(53)	9.6	(52)	9.3	(44)
ACT	1.6	(10)	2.6	(14)	1.5	(7)
Nth Qld	1.1	(7)	1.1	(6)	0.8	(4)
Tas	3.1	(19)	3.5	(19)	3.8	(18)
Undergraduate degree						
Older universities	66.4	(345)	65.4	(299)	62.0	(241)
Newer universities	33.6	(175)	34.6	(158)	38.0	(148)

The study findings for stage of training, workload and trainees in the longitudinal cohort, feature in tables of statistically significant results at the level of $p < .05$. The lower the significance level, the more confident one can be that observed differences reflect real differences in the population (de Vaus, 2002). For time-based variables such as stage of training, the focus was on improvement in self-assessed competence, rather than absolute scores. Appendices B 1 to B 37 show the remaining results. Results on the variables of age and undergraduate education feature only in the text. Unless otherwise stated, the results pertain to years one to three of the surveys.

Domain One: Communication Skills and the Patient-Doctor Relationship

Questions about providing continuity of care using a whole-person approach; communicating with patients, and counselling effectively were included in *Domain one*. In year three, trainees reported a mean score of 4.5 (SD 0.6) in communicating with patients on a scale ranging from 1–5 (1=Strongly disagree and 5=Strongly agree) and 4.4 (SD 0.0) for providing continuity of care using a whole-person approach. They viewed themselves as less competent in counselling patients effectively (mean 4.0, SD 0.8).

Differences between trainees at various stages of training (using ANOVA) provide a more differentiated picture of self-assessment in *Domain one* than do simple frequencies. Trainees viewed themselves as improving most in providing continuity of care using a whole-person approach ($p < .001$). The variations between groups at different stages were quite marked. Former trainees over the three years scored themselves an average of 13 points higher (on a one hundred-point scale) than basic trainees. A Scheffé post hoc test showed that differences between groups for each year, were generally statistically significant between hospital trainees and all other stages of training (Table 3). Although former

trainees also viewed themselves as more competent in communicating effectively with patients, the differences between them and their hospital counterparts were substantially less for years one and two ($p < .05$).

Groups t tests in year two showed that gender differences (although very small) were statistically significant; but only in communicating with patients effectively ($p < .05$). Females viewed themselves as slightly better communicators with patients than males, as shown in Appendix B 1. Similarly, in year three, differences between older and younger trainees were small, but statistically significant. Trainees 30 years and over viewed themselves as slightly more competent than those under 30 years in providing continuity of care using a whole-person approach and counselling patients effectively ($p < .05$).

The provision of continuity of care using a whole-person approach was the only area in *Domain one* in which the number of patients seen by trainees made any statistically significant difference. ANOVA showed that in years one and two trainees seeing an average of between 121 and 160 patients per week assessed themselves to be most competent in the area ($p < .01$ or better). Table 4 shows the results. A Scheffé post hoc test shows that in year two of the survey, these differences were statistically significant between trainees seeing between 81 and 120 patients compared with 121 to 160 patients, and 161 and above.

In providing continuity of care using a whole-person approach, a groups t test shows that rural trainees viewed themselves as slightly more competent than urban ones ($p < .05$). See Appendix B 2.

Table 3: Stage of training by Domain one

	Scale scores					Statistical test		
	Mean* (Standard deviation)					ANOVA F	df	Sig.
	H†	B‡	A§	S	F¶			
Provide continuity of care using a whole-person approach								
1999	86.4 (21.1)	94.2 (14.8)	97.1 (10.7)	98.6 (6.6)	99.0 (6.8)	19.9	4	<i>p</i> <.001
2000	85.9 (20.2)	95.2 (12.6)	99.0 (5.8)	98.3 (7.4)	98.7 (6.6)	22.3	4	<i>p</i> <.001
2001	85.2 (22.1)	95.5 (11.6)	98.2 (7.6)	98.4 (7.0)	99.7 (3.4)	25.6	4	<i>p</i> <.001
Communicate effectively with patients								
1999	97.4 (9.9)	98.3 (8.7)	98.2 (7.0)	100.0 (0.0)	99.4 (5.8)	2.6	4	<i>p</i> <.05
2001	96.9 (9.8)	100.0 (00.0)	98.8 (6.2)	99.7 (2.9)	99.8 (2.4)	5.1	4	<i>p</i> <.001
Counsel patients effectively								
1999	81.0 (24.6)	82.5 (23.8)	88.2 (18.2)	89.4 (19.2)	91.8 (17.0)	6.6	4	<i>p</i> <.001
2000	86.0 (21.4)	79.4 (20.9)	91.3 (17.0)	89.5 (17.5)	91.8 (16.8)	8.2	4	<i>p</i> <.001
2001	88.9 (18.3)	84.7 (23.0)	85.5 (22.0)	92.5 (15.7)	94.2 (14.0)	4.9	4	<i>p</i> <.01

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients ||≤161 patients

Table 4: Workload by Domain one

	Scale scores Mean* (Standard deviation)					Statistical test		
	1†	2‡	3§	4¶	5	ANOVA F	df	Sig.
Provide continuity of care using a whole-person approach								
1999	91.8 (16.6)	92.8 (16.7)	97.8 (9.4)	98.2 (9.0)	95.8 (12.9)	6.4	4	$p<.001$
2000	93.0 (17.1)	95.7 (11.9)	98.1 (7.7)	98.2 (7.5)	91.2 (17.0)	6.8	4	$p<.01$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients || ≤161 patients

Longitudinal Results on Domain One

A sample of 134 trainees participated in all of the surveys. Repeated measures ANOVA showed that trainees viewed themselves as improving most in counselling patients effectively from a mean score of 81.7 (*SD* 24.1) to 92.5 (*SD* 16.7). This finding was statistically significant (at the level of $p<.01$).

Trainees viewed themselves as improving least in communicating with patients effectively, from a mean score of 97.0 (*SD* 11.2) to 99.8 (*SD* 2.9). These findings were statistically significant (at the level of $p<.05$). These differences between the results of those trainees from the longitudinal cohort, and those from the general survey population, suggests some value to this approach that follows individuals. See Table 5.

Table 5: Longitudinal findings for Domain one

	Scale scores Mean* (Standard deviation)			Statistical test		
	1999	2000	2001	ANOVA F	df	Sig.
Provide continuity of care using a whole-person approach	90.2 (18.2)	98.0 (8.0)	99.2 (5.0)	15.9	2	$p<.01$
Communicate with patients effectively	97.0 (11.2)	99.3 (5.0)	99.8 (2.9)	3.8	2	$p<.05$
Counsel patients effectively	81.7 (24.1)	88.0 (18.5)	92.5 (16.7)	11.5	2	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Domain Two: Applied Professional Knowledge and Skills

Questions about clinical areas such as the application of principles of primary medical care, management of a wide range of health problems and medical conditions and critical evaluation of relevant information were asked under *Domain two* items. This domain also featured questions relating to the medical team, such as making clear, concise and appropriate referrals, and the ordering of cost-effective investigations.

In year three, trainees obtained a mean score of 4.2 for the application of principles of primary health care (*SD* 0.6), and 4.1 for the management of a wide range of health problems and medical conditions (*SD* 0.7) and making concise, clear and appropriate referrals (*SD* 0.6). They viewed themselves as less competent in critically evaluating relevant information (mean 3.9, *SD* 0.7) and the ordering of cost-effective investigations (mean 3.9, *SD* 0.7). The surveys from years one and two mirrored these results.

Trainees at different stages of the program in all years reported statistically significant differences in *Domain two* items that were statistically significant ($p < .01$ or better) as shown in Tables 6 and 7. Trainees in later stages of the program viewed themselves as having improved most in the management of a wide range of health and medical problems, and the ordering of cost-effective investigations. In both instances, former trainees rated themselves between 15 and 18 points higher than hospital trainees. A Scheffé post hoc test showed that differences were statistically significant between hospital compared with all other stages of training.

Gender only made a difference to self-assessed competence scores in ordering cost-effective investigations ($p < .05$). Males in years one and two viewed themselves as slightly more competent than females in this area. See Appendix B 3.

Older trainees viewed themselves as improving somewhat more than younger ones in the management of a wide range of health and medical conditions ($p < .01$). In year two of the three surveys, this was followed by the application of principles of primary health care ($p < .01$), critical evaluation of relevant information ($p < .05$), and making concise, clear and appropriate referrals ($p < .01$).

Trainees in year three who saw between 121 and 160 patients viewed themselves as improving most in the critical evaluation of relevant information ($p < .01$). Trainees (in all years) in the management of a wide range of medical conditions ($p < .05$) followed this. In this case, differences were statistically significant between trainees seeing between 1 and 40 patients, 41 and 80 patients, 81 and 120 patients, 121 and 160 patients, and 161 and above patients (Tables 8 and 9).

Table 6: Stage of training by Domain two

	Scale scores Mean* (Standard deviation)					Statistical test		
	H†	B‡	A§	S	F¶	ANOVA F	df	Sig.
Management of a wide range of health and medical problems								
1999	79.8 (23.1)	89.5 (19.5)	97.5 (10.2)	98.1 (7.8)	97.8 (9.1)	34.4	4	$p<.001$
2000	83.1 (22.7)	91.4 (16.0)	96.9 (2.8)	96.3 (10.5)	99.1 (5.4)	21.6	4	$p<.001$
2001	81.5 (23.0)	91.0 (16.9)	94.5 (12.4)	97.4 (9.9)	99.5 (4.1)	27.8	4	$p<.001$
Apply principles of primary health care								
1999	86.5 (20.6)	93.9 (14.4)	97.8 (8.3)	97.3 (10.1)	97.8 (9.8)	15.9	4	$p<.001$
2000	88.8 (18.2)	93.0 (13.6)	97.9 (8.1)	97.5 (9.6)	98.9 (6.0)	14.0	4	$p<.001$
2001	88.9 (19.4)	97.2 (9.3)	98.8 (6.3)	95.9 (11.8)	99.0 (5.8)	10.8	4	$p<.001$
Order cost-effective investigations								
1999	81.9 (23.8)	83.2 (21.8)	86.5 (19.2)	90.0 (16.5)	91.0 (17.0)	5.5	4	$p<.001$
2000	75.2 (26.7)	84.8 (21.7)	89.2 (16.8)	90.9 (16.5)	92.7 (16.8)	12.8	4	$p<.001$
2001	78.4 (24.4)	85.6 (18.5)	85.5 (19.0)	91.5 (16.3)	93.5 (14.5)	9.9	4	$p<.01$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ||Subsequent trainees ¶Former trainees

Table 7: Stage of training by referrals

	Scale scores Mean* (Standard deviation)					Statistical test		
	H†	B‡	A§	S	F¶	ANOVA F	df	Sig.
Make concise, clear and appropriate referrals								
1999	91.2 (17.2)	90.2 (18.2)	95.3 (12.7)	95.2 (13.3)	97.8 (9.1)	6.7	4	$p<.001$
2000	88.5 (20.2)	92.1 (16.4)	97.4 (9.0)	97.5 (8.7)	96.7 (11.4)	8.6	4	$p<.001$
2001	89.5 (18.1)	92.8 (13.9)	93.9 (14.5)	97.4 (9.9)	99.0 (5.8)	10.2	4	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ||Subsequent trainees ¶Former trainees

Table 8: Workload by Domain two

	Scale scores Mean (Standard deviation)					Statistical test		
	1†	2‡	3§	4¶	5	ANOVA F	df	Sig.
Management of a wide range of health problems and medical conditions								
1999	90.0 (17.4)	91.6 (17.4)	95.5 (13.8)	95.4 (14.3)	93.9 (12.5)	3.3	4	$p<.05$
2000	88.1 (20.3)	94.4 (13.8)	95.2 (13.6)	97.8 (8.3)	91.6 (17.5)	4.2	4	$p<.01$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients ||≤161 patients

Table 9: Workload by further items in Domain two

	Scale scores Mean (Standard deviation)					Statistical test		
	1†	2‡	3§	4¶	5	ANOVA F	df	Sig.
2001	92.9 (9.6)	96.7 (11.8)	97.3 (9.9)	99.5 (4.0)	99.3 (4.9)	3.3	4	$p<.05$
Apply principles of primary health care								
2000	95.5 (13.6)	94.4 (13.1)	93.9 (10.3)	97.8 (9.9)	93.3 (13.4)	2.4	4	$p<.05$
Critically evaluate relevant information								
2001	86.3 (18.8)	93.8 (15.2)	89.3 (16.5)	92.3 (16.3)	96.4 (10.5)	3.9	4	$p<.01$
Order cost-effective investigations								
1999	85.8 (21.8)	83.9 (21.1)	86.9 (19.7)	92.6 (15.6)	88.6 (18.0)	2.9	4	$p<.05$
Make concise, clear and appropriate referrals								
2000	93.3 (16.8)	94.9 (13.3)	95.7 (12.2)	98.2 (7.5)	90.5 (18.6)	3.8	4	$p<.01$
2001	91.1 (17.4)	97.0 (10.5)	96.4 (11.0)	99.0 (5.6)	98.6 (6.9)	4.8	4	$p<.01$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients ||≤161 patients

Longitudinal Results on Domain Two

A group of trainees participating in each of the three surveys provided comparisons on *Domain two* items using repeated measures ANOVA. Trainees viewed themselves as improving most in *Domain two* items in the management of a wide range of health and medical problems. In this area, they improved from a mean score of 88.0 (*SD* 20.2) in year one, to a mean score of 98.0 (*SD* 9.0) in year three ($p<.001$). They viewed themselves as improving least in the critical evaluation of relevant information, reporting a mean score of 88.1 (*SD* 18.9) in year one, to 93.4 (*SD* 13.3) in year two ($p<.05$). See Table 10.

Table 10: Longitudinal findings for Domain two

	Scale scores Mean* (Standard deviation)			Statistical test		
	1999	2000	2001	ANOVA F	df	Sig.
Management of a wide range of health and medical problems	88.0 (20.2)	95.5 (12.2)	98.0 (9.0)	15.8	2	$p<.001$
Apply principles of primary health care	90.5 (18.6)	95.3 (12.4)	97.5 (9.7)	8.0	2	$p<.01$
Critically evaluate relevant information	88.1 (18.9)	92.2 (14.8)	93.4 (13.3)	4.4	2	$p<.05$
Order cost-effective investigations	83.0 (22.3)	89.0 (19.1)	93.5 (14.5)	15.1	2	$p<.001$
Make concise, clear and appropriate referrals	91.8 (16.6)	95.8 (12.5)	98.5 (6.9)	11.4	2	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Domain Three: Population Health and the Context of General Practice

Domain three contained questions about making clinical decisions based on the best available evidence, functioning effectively in the health system, and offering integrated care. It also included understanding the relationship between patterns and prevalence of disease and general practice, and undertaking a public health role.

In year three, trainees reported that they were most competent in functioning effectively in the health system and offering integrated care (mean 4.1, *SD* 0.7). They also reported that they were competent in making clinical decisions based on the best available evidence (mean 4.0, *SD* 0.7) and understanding the relationship between patterns and prevalence of disease and general practice (mean 3.9, *SD* 0.7).

Trainees rated undertaking a public health role as the lowest of all *Domain three* items. In year three, trainees reported a mean score of 3.8 (*SD* 0.8).

Trainees in later stages of training viewed themselves as most competent in making clinical decisions based on the best available evidence ($p < .01$), and these differences were statistically significant between hospital compared to subsequent and former trainees. Findings were generally consistent across years one, two and three (Scheffé post hoc test).

Trainees viewed themselves as improving most in understanding the relationship between patterns and prevalence of disease and general practice ($p < .01$). See Tables 11 and 12. Trainees in later stages assessed themselves some 15 points higher than hospital trainees. These differences were statistically significant between hospital and advanced trainees and subsequent and former trainees for years two and three (Scheffé post hoc test).

Table 11: Stage of training by Domain three

	Scale scores Mean* (Standard deviation)					Statistical test		
	H†	B‡	A§	S	F¶	ANOVA F	df	Sig.
Make clinical decisions based on the best available evidence								
1999	90.5 (16.3)	91.9 (15.7)	94.6 (13.3)	94.6 (14.4)	95.8 (13.8)	5.5	4	$p<.05$
2000	86.6 (20.6)	91.4 (17.3)	91.3 (17.0)	95.1 (12.5)	94.9 (13.8)	5.0	4	$p<.01$
2001	85.8 (20.1)	93.7 (13.2)	90.9 (16.3)	94.8 (12.8)	96.9 (10.8)	8.0	4	$p<.001$
Function effectively in the health system								
1999	86.3 (19.7)	91.0 (16.7)	94.1 (12.8)	94.6 (13.1)	94.6 (14.3)	6.5	4	$p<.001$
2000	85.1 (22.0)	94.0 (13.7)	94.9 (12.1)	96.3 (11.3)	94.4 (13.6)	8.8	4	$p<.001$
2001	88.9 (20.5)	93.7 (13.2)	93.9 (12.8)	95.1 (12.6)	95.4 (13.4)	2.4	4	$p<.05$
Offer integrated care								
1999	81.7 (22.6)	84.9 (20.8)	92.7 (13.9)	94.3 (12.6)	96.0 (12.1)	17.6	4	$p<.001$
2000	82.9 (20.9)	87.1 (17.6)	92.7 (13.9)	94.0 (14.7)	93.1 (15.1)	8.4	4	$p<.001$
2001	88.9 (20.5)	93.7 (13.2)	93.9 (12.8)	95.1 (12.6)	95.4 (13.4)	2.4	4	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice:
1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ||Subsequent trainees ¶Former trainees

Table 12: Stage of training by public health items

	Scale scores Mean* (Standard deviation)					Statistical test		
	H†	B‡	A§	S	F¶	ANOVA F	df	Sig.
Understand the relationship between patterns and prevalence of disease and general practice								
2000	76.7 (22.9)	84.4 (19.1)	92.3 (16.4)	91.6 (15.1)	92.0 (16.7)	13.7	4	$p<.001$
2001	79.0 (22.7)	86.5 (16.6)	92.1 (14.3)	94.3 (13.9)	93.6 (14.0)	11.9	4	$p<.001$
Undertake a public health role								
1999	74.0 (25.0)	79.7 (21.9)	83.0 (20.7)	84.8 (22.8)	86.1 (20.2)	6.1	4	$p<.001$
2000	76.5 (25.6)	81.1 (19.6)	90.8 (18.2)	87.9 (19.8)	82.4 (22.7)	6.0	4	$p<.001$
2001	76.7 (22.3)	88.3 (16.1)	85.5 (21.0)	87.6 (18.7)	88.4 (19.2)	4.0	4	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ||Subsequent trainees ¶Former trainees

The only item in which gender made any difference to ratings for *Domain three* was in undertaking a public health role. Females viewed themselves as slightly more competent than males in year three ($p<.05$). See Appendix B 4.

Older trainees in years one and two, viewed themselves as slightly more competent than younger ones in functioning effectively in the health system ($p<.05$). Ratings for offering integrated care ($p<.05$) and understanding the relationship between patterns and prevalence of disease and general practice ($p<.01$) differed in all years. Undertaking a public health role had small

statistically significant age differences for year three. Trainees from more recently established universities viewed themselves as more competent than those from more established universities in making clinical decisions based on the best available evidence ($p<.05$).

Table 13: Longitudinal findings for Domain three

	Scale scores Mean* (Standard deviation)			Statistical test		
	1999	2000	2001	ANOVA F	df	Sig.
Make clinical decisions based on the best available evidence	91.6 (15.6)	94.6 (13.6)	96.3 (11.3)	4.1	2	$p<.05$
Offer integrated care	85.4 (21.1)	90.7 (17.1)	96.7 (10.0)	19.6	2	$p<.001$
Function effectively in the health system	91.0 (16.5)	94.7 (12.9)	96.0 (12.4)	5.1	2	$p<.01$
Understand the relationship between patterns and prevalence of disease and general practice	83.7 (22.3)	90.2 (17.3)	95.0 (12.6)	16.8	2	$p<.001$
Undertake a public health role	77.3 (23.0)	84.6 (22.0)	88.9 (17.3)	18.7	2	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Longitudinal Results on Domain Three

The group of 134 trainees participated in each of the surveys. In undertaking a public health role they viewed themselves as improving most, from a mean score in year two of 77.3 (SD 23.0), to 88.9 (SD 17.3) in year three ($p<.001$). This represented a difference of 11 points from years one to three. Self-assessed improvements for functioning effectively in the health system and making clinical decisions based on the best available evidence were considerably smaller, around four points. See Table 13.

Domain Four: Professional and Ethical Role

This domain encompassed questions about providing support and advice on health and illness in an ethical manner, engaging in ongoing, self-directed learning and professional development, and establishing professional support networks.

In year three, trainees reported that they were most competent in providing support and advice in health and illness in an ethical manner (mean 4.4, *SD* 0.6), and least competent in establishing professional networks (mean 3.9, *SD* 0.8).

In establishing professional support networks, hospital trainees viewed themselves as improving most ($p < .01$). See Table 14. These differences were statistically significant between hospital and basic term trainees and all other groups (Scheffé post hoc test). Engaging in ongoing, self-directed learning and professional development followed ($p < .01$). In this area, differences were statistically significant between all stages of training with the exception of basic trainees (in years one and two), and hospital and former trainees (in year three). Although trainees in subsequent general practice experience viewed themselves as much more competent in providing support and advice in an ethical manner, only differences between hospital and subsequent and former trainees were statistically significant in year three (Scheffé post hoc test).

Older trainees viewed themselves as more competent than younger trainees in year one, in establishing professional networks ($p < .01$). The same applied to trainees in year two ($p < .05$). Older trainees in year one also viewed themselves as more competent than their younger peers in engaging in self-directed learning and professional development ($p < .01$).

Table 14: Stage of training by Domain four

	Scale scores Mean* (Standard deviation)					Statistical test		
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
Provide support and advice on health and illness in an ethical manner								
1999	99.1 (17.3)	97.1 (1.0)	97.8 (0.9)	98.4 (8.4)	99.2 (6.3)	8.9	4	<i>p</i> <.001
2000	95.0 (12.0)	95.8 (11.1)	99.0 (5.8)	99.0 (7.0)	98.0 (10.4)	3.5	4	<i>p</i> <.01
2001	93.8 (16.0)	98.2 (7.6)	97.6 (8.7)	98.5 (7.0)	99.0 (5.8)	4.1	4	<i>p</i> <.01
Establish own professional support networks								
1999	76.6 (25.0)	83.5 (19.4)	89.9 (17.6)	89.2 (19.8)	89.2 (19.1)	6.9	4	<i>p</i> <.001
2000	75.9 (24.7)	81.3 (21.6)	91.8 (15.6)	92.2 (15.8)	89.3 (18.3)	13.5	4	<i>p</i> <.01
2001	82.7 (21.2)	84.7 (18.6)	87.9 (17.4)	93.0 (16.0)	89.7 (19.4)	3.7	4	<i>p</i> <.01
Engage in ongoing, self-directed learning and professional development								
1999	89.5 (19.4)	91.3 (17.2)	95.7 (11.3)	95.6 (12.8)	95.1 (13.5)	4.2	4	<i>p</i> <.01
2000	88.8 (18.2)	92.7 (17.3)	97.9 (8.1)	96.8 (11.4)	96.7 (12.1)	7.0	4	<i>p</i> <.01
2001	87.0 (18.8)	95.5 (14.0)	92.7 (13.9)	96.1 (12.2)	95.0 (14.1)	4.4	4	<i>p</i> <.01

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ||Subsequent trainees ¶Former trainees

In engaging in ongoing, self-directed learning, there were statistically significant differences between trainees in year three, with different numbers of patients ($p<.05$). See Table 15. Those trainees with between 121 and 160 patients viewed themselves as most competent in this area compared to other groups; however, these differences were statistically significant between trainees seeing between Trainees in year one, from more recently established universities, viewed themselves as more competent in engaging in ongoing, self-directed learning and professional development ($p<.05$).

Differences between trainees from different localities were in only one area. Rural trainees in year two viewed themselves as more competent than urban trainees, in engaging in ongoing, self-directed learning and professional development ($p<.05$). See Appendix B 5.

Table 15: Workload by Domain four

	Scale scores Mean (Standard deviation)					Statistical test		
	1†	2‡	3§	4¶	5	ANOVA F	df	Sig.
Establish own professional support networks								
1999	84.5 (21.8)	81.9 (22.4)	88.2 (19.2)	91.7 (17.7)	88.8 (19.7)	3.8	4	$p<.01$
2000	81.5 (24.2)	86.1 (22.2)	88.0 (18.1)	91.7 (15.5)	83.7 (21.5)	2.8	4	$p<.05$
Engage in ongoing, self-directed learning								
2001	88.7 (19.4)	96.4 (12.9)	94.8 (12.2)	96.1 (12.2)	94.9 (14.0)	3.3	4	$p<.05$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients || ≤161 patients

Longitudinal Results on Domain Four

Establishing professional networks was an area in which trainees in the longitudinal cohort viewed themselves as improving over time from a mean score of 82.3 (*SD* 22.3) to 94.0 (*SD* 17.8). For engaging in ongoing, self-directed learning and professional development, trainees reported a small improvement from a mean score of 91.3 (*SD* 16.8) to 97.0 (*SD* 11.8). Each of the *Domain four* items for the longitudinal sample of trainees was statistically significant at the level of $p<.01$. See Table 16.

Table 16: Longitudinal findings for Domain four

	Scale scores Mean* (Standard deviation)			Statistical test		
	1999	2000	2001	ANOVA F	df	Sig.
Establish own professional support networks	82.3 (22.3)	90.0 (17.8)	94.0 (14.1)	15.5	2	$p<.001$
Engage in ongoing, self-directed learning and professional development	91.3 (16.8)	96.5 (11.8)	97.0 (10.4)	5.9	2	$p<.01$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Domain Five: Organizational and Legal Dimensions

Domain five included questions about applying relevant legal principles to general practice, effectively maintaining and using patient records and implementing effective practice management systems.

In year three, trainees viewed themselves as most competent in effectively managing and using patient records (mean 4.1, *SD* 0.7) and least competent in effectively implementing practice management systems (mean 3.0, *SD* 0.9).

Trainees in later stages of training viewed that they were more competent than hospital year trainees for all *Domain five* items ($p < .01$ or better). See Table 17. ANOVA revealed that trainees viewed that they had improved most for applying relevant legal principles to practice. In this area there were substantial, statistically significant differences of at least 17 points between trainees at hospital and later stages of training ($p < .01$). Differences between groups on this item were statistically significant between hospital and basic trainees, and advanced, subsequent and former trainees (Scheffé post hoc test). Implementing effective practice management systems was the next area of greatest improvement with statistically significant differences between trainees at different stages ($p < .01$). Generally, differences were statistically significant between hospital and former trainees (Scheffé post hoc test).

In implementing effective practice management systems in year one males viewed themselves as slightly more competent than females; whereas for effectively maintaining and using patient records, females viewed themselves as slightly more competent ($p < .05$). See Appendix B 6.

Older trainees in years two and three viewed themselves as more competent than younger trainees in implementing relevant practice management systems and applying relevant legal principles to general practice ($p < .01$ or better).

Trainees seeing between 81 and 120 patients per week reported the highest mean scores in effectively maintaining and using patient records ($p < .01$). At the point of 161 patients and above, trainees' self-assessed competence ratings decreased. See Table 18. In year two, these differences were statistically significant between trainees seeing over 161 patients compared with those seeing between 41 and 80 patients, and between 81 and 120 patients (Scheffé post hoc test). For applying relevant legal principles to general practice, trainees with between 121 and 160 patients reported the highest mean scores ($p < .05$). In year two, these differences

were statistically significant between trainees seeing 161 patients and over, compared to those seeing between 41 and 80 patients, and between 81 to 120 patients (Scheffé post hoc test).

Table 17: Stage of training by Domain five

	Scale scores Mean* (Standard deviation)					Statistical test		
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
Implement effective practice management systems								
1999	58.5 (25.3)	64.0 (26.4)	66.7 (23.7)	65.9 (24.7)	70.3 (25.6)	3.9	4	$p<.01$
2000	58.8 (24.5)	59.6 (23.1)	72.8 (23.5)	71.1 (25.0)	72.9 (24.6)	8.8	4	$p<.001$
2001	60.5 (23.4)	67.6 (24.2)	67.9 (26.4)	73.9 (23.5)	76.1 (23.9)	5.3	4	$p<.01$
Apply relevant legal principles to own general practice								
1999	63.5 (25.5)	74.3 (25.5)	75.6 (23.6)	80.0 (22.1)	80.9 (22.7)	10.7	4	$p<.001$
2000	62.4 (24.6)	68.9 (24.1)	85.6 (18.6)	83.5 (21.5)	83.3 (23.1)	20.2	4	$p<.001$
2001	61.7 (21.9)	69.4 (22.7)	76.4 (22.8)	83.6 (20.1)	87.0 (19.8)	19.8	4	$p<.01$
Effectively maintain and use patient records								
1999	78.8 (24.7)	92.4 (16.6)	96.4 (10.4)	96.8 (10.8)	95.8 (12.8)	26.8	4	$p<.001$
2000	82.2 (23.8)	94.3 (12.6)	97.9 (8.1)	97.0 (10.4)	96.2 (10.6)	20.1	4	$p<.001$
2001	88.3 (19.6)	94.6 (14.7)	93.9 (14.5)	96.4 (11.2)	96.9 (10.2)	5.3	4	$p<.01$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ||Subsequent trainees ¶Former trainees

For implementing effective practice management systems, trainees with between 121 and 160 patients reported the highest mean scores in year two ($p<.05$).

Trainees in this area viewed that they have improved most. Differences were statistically significant between trainees seeing between 81 and 120 patients and between 121 and 160 patients (Scheffé post hoc test).

Table 18: Workload by Domain five

	Scale scores Mean* (Standard deviation)					Statistical test		
	1†	2‡	3§	4¶	5	ANOVA F	df	Sig.
Effectively maintain and use patient records								
1999	87.0 (23.2)	92.6 (16.8)	95.2 (13.0)	92.6 (16.3)	95.2 (14.3)	3.6	4	$p<.01$
2000	89.9 (20.0)	95.5 (12.9)	96.4 (11.4)	95.6 (11.4)	88.1 (19.4)	6.9	4	$p<.01$
Implement effective practice management systems								
1999	60.0 (27.2)	63.7 (24.0)	64.7 (25.2)	73.5 (24.4)	73.5 (26.9)	4.3	4	$p<.01$
2000	64.3 (24.6)	64.4 (25.6)	67.2 (24.6)	76.0 (24.2)	68.1 (24.8)	2.9	4	$p<.05$
2001	64.9 (23.3)	74.4 (24.1)	74.1 (24.1)	76.8 (25.8)	76.8 (24.2)	2.6	4	$p<.05$
Apply relevant legal principles to own general practice								
2000	76.0 (24.5)	74.0 (25.5)	77.9 (23.7)	85.5 (19.9)	75.8 (25.5)	3.0	4	$p<.05$
2001	74.4 (22.9)	81.5 (23.2)	81.0 (22.0)	86.0 (21.0)	84.8 (19.5)	2.4	4	$p<.05$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients || ≤161 patients

Care of Particular Patient Groups

Trainees answered a group of questions about the care of neonates, infants, children, adolescents, women, men, older and indigenous people, and about those from non-English speaking backgrounds, in rural and remote areas, with mental illness, disabilities, or with chronic conditions.

Trainees viewed themselves as most competent in the care of women, older people, children, men, adolescents, people with chronic conditions and infants. The mean scores for these items ranged from 3.9 to 4.3. Trainees viewed themselves as least competent in the care of neonates, indigenous people, those with mental illness or disabilities, and from non-English speaking backgrounds and from remote, rural and remote communities. The mean score for these groups ranged from 3.2 to 3.7. Some improvement in these areas was evident over the three data collection periods of the study.

Trainees in later stages of training viewed themselves as most competent in the care of neonates, infants, children, adolescents, women, men, indigenous people and those with mental illness, disabilities and chronic conditions ($p < .01$ or better). See Tables 19 to 21. The greatest areas of self-assessed improvement in trainees were in the care of infants followed by neonates, children and people with mental illness. In the care of infants and children for trainees in years one and two, differences were statistically significant between hospital and all other stages of training (Scheffé post hoc test).

Males and females had some statistically significant differences in self-assessments of the care of particular patient groups. Perhaps due to their distinct patient profiles, females in year one viewed themselves as more competent in the care of infants, children ($p < .001$) and neonates ($p < .01$). Not surprisingly, females in all years, viewed themselves as more competent in the care of women ($p < .01$)

Table 19: Stage of training by particular patient groups

	Scale scores					Statistical test		
	Mean* (Standard deviation)					ANOVA F	df	Sig.
	H†	B‡	A§	S	F¶			
Neonates								
1999	54.9 (27.2)	66.4 (28.0)	73.6 (27.0)	78.2 (26.6)	79.5 (25.2)	18.2	4	$p<.01$
2000	59.5 (28.3)	65.1 (26.7)	75.4 (25.2)	74.3 (26.3)	79.2 (26.4)	9.8	4	$p<.001$
2001	60.5 (29.0)	76.6 (22.0)	70.4 (28.0)	79.1 (23.6)	79.5 (25.5)	7.1	4	$p<.001$
Infants								
1999	67.8 (28.8)	86.8 (21.6)	92.0 (16.7)	95.2 (13.2)	97.4 (16.3)	50.1	4	$p<.001$
2000	76.5 (27.3)	87.6 (20.8)	93.3 (15.8)	93.9 (13.6)	95.5 (13.2)	18.2	4	$p<.001$
2001	77.2 (26.6)	95.5 (11.6)	90.1 (19.0)	94.1 (15.3)	97.7 (9.1)	20.1	4	$p<.001$
Children								
1999	81.3 (24.2)	94.4 (14.7)	98.2 (7.6)	98.4 (7.2)	99.0 (6.8)	36.7	4	$p<.001$
2000	84.4 (23.1)	97.1 (9.4)	97.4 (10.7)	98.3 (7.4)	98.4 (9.8)	20.4	4	$p<.001$
2001	86.4 (21.0)	98.2 (7.6)	96.7 (9.8)	99.0 (5.8)	99.3 (4.8)	22.2	4	$p<.001$
Adolescents								
1999	85.6 (21.2)	90.1 (17.7)	95.3 (12.7)	96.8 (9.9)	94.6 (13.3)	10.6	4	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ||Subsequent trainees ¶Former trainees

Table 20: Stage of training by women to indigenous people

	Scale scores					Statistical test		
	Mean* (Standard deviation)					ANOVA F	df	Sig.
Women								
1999	89.9 (19.3)	95.0 (11.9)	97.1 (11.7)	98.7 (6.6)	98.6 (8.5)	11.3	4	$p<.001$
2000	90.5 (18.2)	97.8 (9.5)	98.5 (7.0)	98.0 (7.9)	99.1 (5.4)	11.6	4	$p<.001$
2001	90.7 (16.4)	92.8 (16.0)	97.5 (8.8)	97.7 (8.5)	97.4 (10.2)	5.4	4	$p<.001$
Men								
1999	92.5 (18.2)	93.3 (13.4)	97.1 (9.4)	95.2 (13.9)	96.6 (11.3)	2.6	4	$p<.05$
2000	91.3 (15.6)	94.0 (15.2)	94.9 (12.1)	96.6 (10.9)	97.1 (10.9)	3.5	4	$p<.01$
2001	90.7 (16.4)	92.8 (16.0)	97.5 (8.8)	97.7 (8.5)	97.4 (10.2)	5.4	4	$p<.001$
Indigenous people								
1999	69.3 (25.6)	76.5 (24.8)	84.4 (20.6)	82.8 (21.9)	77.6 (24.3)	6.6	4	$p<.001$
2000	73.5 (26.3)	75.7 (23.0)	79.5 (21.8)	84.4 (21.5)	82.8 (22.1)	4.5	4	$p<.01$
2001	70.4 (26.4)	74.8 (21.4)	79.2 (21.9)	86.0 (19.6)	83.7 (22.1)	6.3	4	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ||Subsequent trainees ¶Former trainees

Table 21: Stage of training by people with mental illness to chronic conditions

	Scale scores Mean* (Standard deviation)					Statistical test		
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
People with mental illness								
1999	72.5 (25.1)	76.9 (23.1)	81.5 (22.8)	84.5 (21.4)	84.0 (20.0)	6.6	4	<i>p</i> <.001
2000	76.5 (24.8)	76.8 (24.1)	85.1 (19.6)	86.5 (20.0)	88.6 (18.5)	7.9	4	<i>p</i> <.001
2001	71.0 (26.7)	83.8 (23.1)	86.9 (22.1)	89.1 (17.8)	85.8 (21.7)	7.4	4	<i>p</i> <.001
People with disabilities								
1999	75.1 (23.3)	79.8 (21.6)	83.7 (21.3)	84.0 (21.8)	86.3 (19.8)	5.4	4	<i>p</i> <.001
2000	78.8 (22.7)	78.1 (23.5)	84.6 (19.6)	87.0 (18.2)	85.9 (19.1)	4.6	4	<i>p</i> <.01
2001	70.4 (23.9)	80.2 (24.2)	77.8 (22.4)	84.5 (19.1)	86.1 (20.8)	6.9	4	<i>p</i> <.001
People with chronic conditions								
1999	82.1 (20.2)	87.4 (19.8)	93.2 (13.6)	95.9 (11.0)	98.3 (8.2)	10.1	4	<i>p</i> <.001
2000	86.7 (20.1)	87.4 (19.0)	93.2 (15.2)	95.5 (12.2)	96.0 (12.0)	10.1	4	<i>p</i> <.01
2001	90.9 (16.6)	91.1 (16.2)	94.4 (15.1)	96.3 (11.9)	97.1 (11.6)	4.9	4	<i>p</i> <.001

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ||Subsequent trainees ¶Former trainees

compared to males who viewed themselves as more competent in the care of men ($p<.01$). Males in years two and three also viewed themselves as more competent than women, in the care of people from remote, rural communities and older people ($p<.01$). See Appendix B 7.

For the care of neonates, infants and children, there were statistically significant differences (at the level of $p<.01$) between older and younger trainees in all survey years. There were also statistically significant differences between trainees in years one and two, for the care of women (at the level of $p<.05$). For the care of people from non-English speaking backgrounds, and those with mental illness or disabilities among trainees in years two and three, older trainees viewed themselves as more competent than younger ones ($p<.05$ or better).

As Tables 22 to 25 demonstrate, trainees in year three with larger numbers of patients viewed themselves improving most in the care of people with disabilities ($p<.001$). The next area of greatest self-assessed improvement, also in year three, was in the care of indigenous people, ($p<.001$). Both of these differences were statistically significant between trainees seeing between 1 and 40 patients and all other groups.

Trainees in year two of the study with between 121 and 160 patients viewed themselves as more competent than other groups of trainees in the care of infants, children and women ($p<.01$). These differences were statistically significant between trainees seeing between 1 and 40 patients, compared with those seeing between 81 and 120 patients, and 121 and 160 patients. For the care of women, the same applied, except that differences were statistically significant between those seeing between 1 and 40 patients, and between 121 and 160 patients.

Table 22: Workload by particular patient groups

	Scale scores					Statistical test		
	Mean* (Standard deviation)					ANOVA F	df	Sig.
Infants								
1999	85.1 (27.0)	86.3 (23.2)	92.0 (17.5)	93.9 (15.5)	93.3 (17.7)	4.4	4	$p<.01$
2000	82.2 (25.2)	90.3 (20.0)	92.9 (15.3)	93.4 (13.4)	86.3 (23.9)	4.6	4	$p<.01$
2001	86.3 (23.6)	96.0 (11.8)	93.2 (16.3)	96.1 (12.3)	97.1 (9.5)	4.9	4	$p<.05$
Neonates								
2001	69.6 (27.2)	72.4 (27.0)	76.2 (25.8)	81.3 (22.6)	82.6 (24.1)	2.9	4	$p<.05$
Children								
1999	88.9 (21.1)	93.9 (15.5)	97.2 (9.8)	97.9 (10.8)	98.1 (7.9)	6.7	4	$p<.01$
2000	89.6 (19.9)	95.9 (12.4)	97.9 (9.4)	99.1 (5.4)	91.8 (19.3)	7.3	4	$p<.01$
2001	93.5 (17.3)	98.2 (7.6)	97.9 (8.1)	99.0 (5.7)	98.6 (6.9)	3.4	4	$p<.01$
Adolescents								
1999	87.4 (20.7)	90.6 (18.2)	95.4 (12.0)	95.3 (12.6)	95.2 (11.8)	5.4	4	$p<.001$
2000	90.4 (16.9)	93.9 (14.2)	96.5 (11.3)	96.5 (11.6)	93.8 (13.9)	3.5	4	$p<.05$
2001	89.9 (20.1)	96.1 (10.8)	96.0 (10.9)	96.1 (12.3)	97.1 (9.5)	3.2	4	$p<.05$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice:
1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients || ≤161 patients

Table 23: Workload by women

	Scale scores Mean* (Standard deviation)					Statistical test		
	1†	2‡	3§	4¶	5	ANOVA F	df	Sig.
Women								
2000	93.3 (16.8)	96.7 (10.8)	98.4 (7.0)	99.6 (3.8)	94.8 (13.9)	4.6	4	$p<.01$
2001	93.5 (16.1)	99.4 (4.5)	97.0 (11.0)	98.0 (9.8)	99.3 (4.9)	3.9	4	$p<.01$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients ||≤161 patient

Trainees in both years two and three viewed themselves as more competent in the care of indigenous people, and patients from rural and remote communities (at a significance level of $p<.05$ or better). See Appendix B 8.

Longitudinal Results on the Care of Particular Patient Groups

Trainees in the longitudinal cohort viewed themselves as improving most in the care of infants, indigenous people and those with mental illness. Predictably, trainees viewed themselves as improving least for areas in which they viewed themselves to be most competent such as the care of older people, women, men, and people with disabilities. With the exception of older people and those with disabilities ($p<.05$), all findings were statistically significant at $p<.01$ or better (Table 26).

Table 24: Workload by men to people with chronic conditions

	Scale scores Mean* (Standard deviation)					Statistical test		
	1†	2‡	3§	4¶	5	ANOVA F	df	Sig.
Men								
1999	93.2 (15.7)	93.2 (15.5)	96.0 (11.3)	97.5 (11.2)	97.1 (9.5)	2.4	4	$p<.05$
2001	94.4 (14.4)	97.3 (9.2)	95.6 (12.5)	99.0 (5.7)	98.6 (6.9)	2.5	4	$p<.05$
Indigenous people								
2001	72.6 (26.3)	86.7 (20.3)	80.2 (21.6)	83.6 (19.6)	88.4 (20.1)	5.3	4	$p<.001$
People from non-English speaking backgrounds								
2001	79.2 (22.5)	90.9 (17.4)	84.6 (21.2)	85.3 (22.6)	89.1 (17.3)	3.7	4	$p<.01$
People from rural remote communities								
1999	78.4 (22.1)	80.7 (22.2)	84.9 (20.8)	87.1 (19.7)	80.9 (25.9)	2.5	4	$p<.05$
2001	79.2 (24.3)	88.2 (20.5)	88.6 (19.5)	88.1 (18.1)	92.0 (16.0)	3.2	4	$p<.05$
People with chronic conditions								
1999	87.0 (19.2)	91.1 (17.9)	93.7 (14.6)	94.6 (14.2)	96.2 (10.8)	3.6	4	$p<.01$
2001	89.9 (17.8)	95.8 (12.9)	94.3 (13.7)	97.5 (8.8)	97.1 (9.5)	3.3	4	$p<.05$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients || ≤161 patients

Table 25: Workload by people with disabilities

	Scale scores Mean* (Standard deviation)					Statistical test		
	1†	2‡	3§	4¶	5	ANOVA F	df	Sig.
People with disabilities								
2001	69.6 (24.0)	87.0 (21.2)	84.4 (20.1)	82.4 (21.9)	88.4 (16.1)	7.7	4	p<.001

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients || ≤161 patients

Care of Patient Presentations

A range of questions elicited from trainees their degree of self-assessed competence in the management of the particular patient presentations represented in the general practice vocational training curriculum. These were areas identified with medical educators during the curriculum development process.

Trainees viewed themselves as most competent in the management of respiratory, gastrointestinal and digestive problems, cardiovascular problems, and paediatrics and preventive medicine. The means scores for these items ranged from 4.0 to 4.2.

Trainees viewed themselves as least competent in the management of major trauma and disasters and HIV/AIDS. The mean scores for these items ranged from 2.8 to 3.2. A relatively large percentage of trainees neither agreed nor disagreed that they were competent in HIV/AIDS management. This is of interest, if only for the reason that trainees might be less willing to concede that this as an area of weakness.

Table 26: Longitudinal findings for particular patient groups

	Scale scores Mean* (Standard deviation)			Statistical test		
	1999	2000	2001	ANOVA F	df	Sig.
Neonates	62.0 (28.9)	70.8 (27.0)	77.8 (25.0)	19.3	2	$p<.001$
Infants	83.2 (23.9)	94.7 (13.6)	96.9 (10.5)	25.3	2	$p<.001$
Children	93.1 (15.1)	99.0 (5.8)	99.7 (2.9)	12.4	2	$p<.001$
Adolescents	90.3 (18.3)	95.7 (12.0)	96.9 (9.7)	7.9	2	$p<.01$
Women	94.9 (12.8)	98.2 (7.6)	98.2 (9.6)	4.8	2	$p<.01$
Men	94.7 (13.6)	95.9 (13.1)	98.0 (8.0)	5.1	2	$p<.01$
Older people	94.4 (15.6)	96.4 11.9	97.5 (8.9)	3.1	2	$p<.05$
Indigenous people	73.7 (25.3)	79.9 (22.7)	86.2 (19.9)	18.2	2	$p<.001$
People from rural remote communities	80.5 (23.9)	87.5 (20.1)	88.9 (20.2)	7.9	2	$p<.01$
People from non-English speaking backgrounds	81.4 (22.8)	85.8 (21.2)	88.1 (18.1)	5.1	2	$p<.01$
People with mental illness	76.8 (23.8)	84.5 (20.8)	88.3 (18.9)	16.8	2	$p<.001$
People with disabilities	79.5 (20.1)	83.1 (22.1)	84.9 (20.0)	3.2	2	$p<.05$
People with chronic conditions	89.8 (17.0)	94.1 (14.6)	97.5 (8.9)	14.4	2	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

One-way ANOVA showed the largest statistically significant differences between trainees in all years at different stages of training in the care of patients with dermatological, and ear, nose and throat problems ($p<.01$). See Appendix B 9. In the case of ear, nose and throat problems, subsequent or former trainees viewed themselves as improving most (Appendices B 10 to B 14). The differences between trainees at the various stages of training were statistically significant between hospital or basic trainees compared to advanced and former trainees (Scheffé post hoc tests). Subsequent trainees also viewed themselves as having improved most in the management of eye problems ($p<.01$), and least for the management of respiratory problems and emergency medicine ($p<.05$). Differences were statistically significant between hospital and basic trainees, compared to later stages of training for most items. The exceptions were the care of patients with HIV/AIDS and the management of emergency medicine. Another pattern prevailed for dermatological, musculoskeletal and gastrointestinal problems. In these areas, differences were statistically significant between hospital trainees compared to subsequent and former trainees (Scheffé post hoc test).

Males and females reported statistically significant differences in relation to their self-assessed competence in the care of particular patient presentations. Researchers have suggested that different patient populations present to male and female trainees. In this context, the gender differences were foreseeable. Males viewed themselves as more competent than females in the management of major trauma and disasters, emergency medicine, and eye, musculoskeletal, neurological and drug and alcohol problems ($p<.01$), as well as cardiovascular problems and emergency medicine ($p<.05$). Females, on the other hand, viewed themselves as more competent in the management of reproductive problems, sexually transmitted diseases, drug and alcohol problems ($p<.01$) and preventive medicine (Appendices B 15 to B 17).

Older trainees in all years viewed themselves as more competent than younger ones in the management of particular patient presentations. These differences were statistically significant at a level of $p < .05$ or better. The care of patients with dermatological problems is particularly noteworthy because older respondents in years two and three surveys reported a mean score of at least 11 points higher than those who were younger ($p < .01$ or better).

The self-assessments of trainees with patient loads of between 121 and 160 patients improved most in the care of ear, nose and throat, gastrointestinal and dermatological problems ($p < .01$), and reproductive, neurological, musculoskeletal and psychological problems ($p < .05$). See Appendices B 18 to B 22. In the management of musculoskeletal and gastrointestinal problems, these differences were statistically significant between those trainees year three seeing between 1 and 40 patients and 121 and 160 patients (Scheffé post hoc test).

In preventive and public health medicine, trainees seeing up to 160 patients agreed that they were more competent than other groups ($p < .01$). The differences between groups in year three were not statistically significant (Scheffé post hoc test). A similar situation prevailed for groups in year two with one exception. For the management of public health medicine, differences were statistically significant between trainees seeing between 41 and 80 patients and between 121 and 160 patients (Scheffé post hoc test). In the same area, trainees in year three, seeing between 81 and 160 patients had a mean score of between five and 15 points higher than those in year two seeing above 161 patients per week.

There were few areas of difference between trainees with various types of undergraduate education. Trainees in year one from more recently established universities agreed that they were slightly less competent in the care of people with eye problems ($p < .01$). The same applied for trainees in year two in the

care of neurological problems ($p<.05$), and trainees in year three for the care of sexually transmitted diseases ($p<.05$). In the case of locality, trainees from remote, rural areas viewed themselves as more competent. See Appendix B 23.

Longitudinal Results on Patient Presentations

The findings on the self-assessed competence of the longitudinal group of trainees completing each of the surveys were compared for patient presentation items using repeated measures ANOVA. Trainees viewed themselves improving most in the management of dermatological, eye, and ear, nose and throat problems, as well as sexually transmitted diseases ($p<.01$ or better). It is noteworthy that over the survey years, there was an improvement in the self-assessed mean score of 24 points in the management of dermatological problems. The smallest perceived improvement occurred with respiratory problems ($p<.05$). See Appendix B 24 and Appendix B 25.

Satisfaction

Trainees in each of the surveys reported they were satisfied with most aspects of training over the three years in a series of questions ranging from 1–5 (1=Very dissatisfied, 5=Very satisfied). They viewed themselves as most satisfied with the range of clinical experiences during training, learning challenges provided, opportunities to learn about communication skills, and role models provided by supervisors. They were similarly satisfied with access to educational material and resources, opportunities to meet with other trainees, level of workload expected, supervision and advice received, information provided, location of training posts and attachments, and external clinical teaching sessions. The mean scores for these items ranged from 3.6 to 3.9.

Trainees reported being less satisfied with the administration of the program, opportunities to learn about population health in the context of general practice, other terms and conditions, remuneration and allowances, and video and

consultation reviews. The mean scores for these items range from 3.1 to 3.5. The unpopularity of the logbook (designed to document training activity) was consistent between the third year and all other years. The mean score ranged from 2.5 (*SD* 0.9) to 2.7 (*SD* 0.9). Similarly, the large percentage of trainees who reported that they viewed themselves as neutral about the usefulness of the logbook and video and consultation reviews suggests they might not be frequently undertaking these activities.

Trainees at different stages of training showed slightly different patterns from year to year of the survey; but the overall picture was consistent as shown in Appendices B 26 to B 31. In years one and two, a similar pattern prevailed with trainees in advanced or subsequent experience reporting higher mean satisfaction scores in all areas ($p < .05$ or better). These differences were generally statistically significant between hospital, advanced, subsequent and former trainees in year one, and hospital and former trainees in year two (Scheffé post hoc test). For the usefulness of the logbook, however, those in the hospital term were most satisfied ($p < .01$) and these differences were statistically significant between hospital and advanced and former trainees (Scheffé post hoc test). This suggests that as trainees progressed, they might have engaged more with the logbook and become less satisfied with it. Another exception to this pattern was satisfaction with remuneration and allowances received. Subsequent trainees in year two were more satisfied than all groups except former trainees. Differences were statistically significant between those in the hospital and basic terms (Scheffé post hoc test).

In year three, subsequent trainees reported statistically significant differences for all satisfaction items, except the usefulness of the logbook and the type of learning challenges provided ($p < .001$). For opportunities to learn communication skills and information sent out to trainees, the differences between individual

stages of training, however, were only statistically significant between hospital and advanced trainees (Scheffé post hoc test).

Male and female trainees reported differences in individual areas of satisfaction that were statistically significant. Females in years one or two of the survey were slightly more satisfied with other terms and conditions of employment, and access to educational materials and resources ($p<.05$). Males, on the other hand, were slightly more satisfied with video and consultation reviews and opportunities to learn about population health ($p<.05$). See Appendix B 32.

Trainees under 30 years of age were more satisfied than older trainees with role models provided by supervisors, other terms and conditions of employment ($p<.01$), supervision, advice, remuneration and allowances received, and learning challenges provided ($p<.05$). The inverse applied to older trainees, who were more satisfied with opportunities to learn about population health, video and consultation reviews, the trainee logbook and the level of workload expected ($p<.05$). These results might suggest that as trainees progressed they became more positively disposed toward educational activities that medical educators deemed to be of value.

Locality made a difference to trainee satisfaction in three areas. Rural and remote trainees in year two were somewhat more satisfied with the range of clinical experiences ($p<.05$) and trainees in years one and two were more satisfied with the location of training posts ($p<.01$). Urban trainees in all years, on the other hand, were more satisfied with their access to educational materials and resources ($p<.05$ or better). These findings reflect that urban trainees viewed themselves as better catered for by the program; but rural trainees were more satisfied with the location of their training posts probably because they have opted to be in a rural area. See Appendix B 33.

Trainees with different workloads held different views on satisfaction as shown in Appendices B 34 and 35. For the range of clinical experiences, those trainees in years one and two seeing between 81 and 120 patients were more satisfied than trainees seeing 121 and over ($p < .05$ or better). The same applied to trainees in year two for external clinical teaching sessions and role models provided by supervisors ($p < .01$), and opportunities to meet with other trainees ($p < .05$). In the case of opportunities to meet with peers and other terms and conditions of employment ($p < .01$), trainees in year two become less satisfied the more patients they saw ($p < .05$). However, the only areas where differences between all groups of trainees in year two with different numbers of patients were statistically significant were role models provided by supervisors and external clinical teaching sessions. In these instances, differences between trainees seeing from 41 to 80 patients, 81 to 120 patients and 161 patients and above were statistically significant (Scheffé post hoc test). In year one, trainees seeing 121 to 160 patients were most satisfied with opportunities to learn about population health in the context of general practice; however, these differences were only statistically significant between those trainees with between one and 40 patients, and 121 and 160 patients (Scheffé post hoc test). In year three, the only areas in which differences between groups of trainees based on workload were statistically significant between all groups, were opportunities to learn communication skills and information sent to trainees (Scheffé post hoc test).

Trainees in the second year with different undergraduate backgrounds reported different views on the usefulness of the logbook. Those from older universities reported being somewhat more satisfied with the usefulness of the logbook; however, both groups were fairly dissatisfied with this item ($p < .01$).

Longitudinal Findings on Satisfaction

Satisfaction data on the group of longitudinal respondents were analysed using repeated measures ANOVA. On all satisfaction items, trainees in the second year were more satisfied. These findings across so many satisfaction items among the same trainees suggest that as trainees progressed, they became more satisfied with aspects of the program. This effect was more marked in the longitudinal cohort of the same trainees participating in all three surveys than the general survey population each year. One explanation for this lies in the progression of stages of training, particularly in the move from hospital to the basic general practice terms. On some satisfaction items, trainees in the third year were also more satisfied. Trainees' satisfaction improved most for the level of workload expected, opportunities to learn about population health, remuneration and allowances received, video and consultation reviews, other terms and conditions of employment, administration of training and the range of clinical experiences offered ($p < .01$ or better). Trainee' satisfaction improved least with external clinical teaching sessions, supervision and advice ($p < .05$), learning challenges provided, and opportunities to learn communication skills ($p < .01$), but these items were already scored very high (Appendices B 36 to B 37).

Satisfaction and Self-assessed Competence

The self-assessment model in Figure 1 on page 94 suggested a possible relationship between satisfaction and self-assessed competence. In other words, the extent to which trainees are satisfied with the program of training might determine their assessments of competence. Block recursive regression was the analytic tool appropriate to exploring this hypothesis. To demonstrate the total causal effects of independent variables upon a dependent variable with block recursive regression, a scale was required to develop measure of satisfaction. Each of the survey years required a different approach to the development of scales for the regression.

For year one, satisfaction with work conditions, program components, opportunities to learn about population health, and teaching and learning outcomes, were the dependent variables on a block recursive multiple regression. For this design, stage of training, gender, age, full- or part-time status, workload, university of graduation, locality and state were the independent variables.

The independent variables in the model explained 32 percent of the variance in a scale measuring five domains of general practice competence. By social science standards, this result is acceptable. The standardised partial correlation coefficient, beta measured between 0 and 1. This showed that having completed training emerged as the strongest predictor in determining how competent trainees view themselves on the five domains of general practice scale (beta=.25) with higher satisfaction representing higher self-assessed competence. The unstandardised regression coefficient showed that a coefficient of 5.50 for the former stage of training indicated that a one-unit increase induced a change of 5.5 points on a scale of 1 to 100.

Satisfaction with opportunities to learn about population health was the second most important predictor of self-assessed competence (beta=.25). A one-unit increase was associated with a change of .25 points on a scale of one to 100. This meant that an improvement of 10 points on the satisfaction with opportunities to learn about population health scale resulted in an improvement of 2.5 percent on the general practice competence scale.

The third predictor of self-assessed competence was satisfaction with administration (beta=.18). A one-unit increase in satisfaction with conditions of work (location of training posts, remuneration and allowances, level of workload, other terms and conditions and administration) resulted in a 0.11 percent increase on the competence scale. An improvement of 10 points on satisfaction with conditions of work resulted in an improvement of 1.1 percent on the general practice competence scale. See Table 27.

Table 27: Predictors of self-assessed competence for year one

Independent variables	Dependent variable five domains of general practice scale*	
	B Unstandardised coefficient	Beta Standardised coefficient
Former trainee stage of training	5.50	0.26
Satisfaction with opportunities to learn about population health	0.11	0.25
Satisfaction with administration	0.10	0.18
Explained variance (R ²)		0.32

*All coefficients significant at the level of $p < .05$.

For year two, using satisfaction with communication and administration as predictors of self-assessed competence, a block recursive multiple regression analysis was performed using stage of training, gender, age, full or part-time status, workload, university of graduation, locality and state as independent variables. The result was that the independent variables in the model explained 41 percent of the variance in a scale measuring five domains of general practice competence. By social science standards, this is very acceptable. Teaching challenges included the range of clinical experiences, learning challenges provided, role models provided by supervisors, and supervision and advice received. The standardised partial correlation coefficient (beta) measured between 0 and 1, showed that teaching challenges emerged as the strongest predictor in

determining how competent trainees viewed themselves on the five domains of general practice scale ($\beta=.30$) with higher satisfaction representing more self-assessed competence. The unstandardised regression coefficient of .24 for teaching challenges indicated that a one-unit increase was associated with a change of .24 points on a scale of 1 to 100. In other words, an improvement of 10 points on the teaching scale resulted in an improvement of 2.40 percent on the five domains of general practice self-assessed competence scale.

Stage of training was the second most important predictor of self-assessed competence ($\beta=.26$). The unstandardised regression coefficient was 1.90. As each term of the training program progressed, self-assessed competence increased by 1.90 percent (on a scale of 1 to 100). In other words, an improvement of 10 points on the teaching scale was associated with an improvement of 19 percent on the five domains of general practice competence scale.

The third predictor of self-assessed competence in year two was satisfaction with opportunities to learn about population health ($\beta=.19$). A one-unit increase in satisfaction with opportunities to learn about population health was associated with a 0.11 percent increase on the self-assessed competence scale. See Table 28.

A block recursive multiple regression analysis was used for the third year of the study. The analyses explained 50 percent of the variance in the scale of self-assessed competence in five domains of general practice. The standardised partial correlation coefficient (β) measured between 0 and 1, showed that stage of training emerged as the strongest predictor in determining how competent trainees view themselves on the five domains of general practice competence scale ($\beta=0.33$). The unstandardised regression coefficient showed that with a coefficient of 2.19 for stage of training, a one-unit increase was associated with a change of 2.2 points on a scale of one to a hundred.

Table 28: Predictors of self-assessed competence for year two

Independent variables	Dependent variable five domains of general practice scale*	
	B Unstandardised coefficient	Beta Standardised coefficient
Satisfaction with the range of clinical experiences, learning challenges, role models, and supervision and advice	0.24	0.30
Stage of training	1.90	0.26
Opportunities to learn about population health	0.11	0.19
Explained variance (R ²)		0.41

*All coefficients significant at the level of $p < .05$.

The next most important predictor of self-assessed competence was satisfaction with communication factors (beta .22). The scale consisted of satisfaction with communication skills, opportunities to meet other trainees, external clinical teaching sessions, and video and consultation reviews. The unstandardised regression coefficient was .12. A one-unit increase in satisfaction with communication factors resulted in a .12 percent increase on the self-assessed competence scale. This meant that an improvement of 10 points on the teaching scale resulted in an improvement of 1.20 percent on the self-assessed competence scale.

Satisfaction with opportunities to learn about population health emerged as the next strongest predictor of self-assessed competence (beta=.20). The unstandardised regression coefficient was .01. A one-unit increase in satisfaction with opportunities to learn about population health resulted in a 0.1 percent increase on the self-assessed competence scale. In other words, an improvement of 10 points in the teaching scale resulted in an improvement of 1.0 percent on the self-assessed competence scale. See Table 29.

Table 29: Predictors of self-assessed competence for year three

Independent variables	Dependent variable five domains of general practice scale*	
	B Unstandardised coefficient	Beta Standardised coefficient
Stage of training	2.19	0.33
Satisfaction with communication factors	0.12	0.22
Opportunities to learn about population health	0.01	0.20
Explained variance (R ²)		0.50

*All coefficients significant at the level of $p < .05$.

In addition to the role of stage of training, satisfaction with opportunities to learn about population health explained an increase on the self-assessed competence scale. This emphasis on population health as a dimension of public health is particularly noteworthy for program developers seeking to address aspects of the curriculum in which trainees might be interested.

The Relationship Between Self-assessments and Examination Results

For years one to three of the survey, the sample sizes for survey respondents who had completed their examinations were $n=94$, 96 and 112. The numbers in the samples for each of the three years were reasonably small due to the post hoc method used to match the trainees. They were included in the analysis if they had completed the examination within two months of receiving each questionnaire.

Self-assessed Competence

In year one, there was a weak, statistically significant relationship between self-assessed competence in undertaking a public health role, and the overall mean examination score ($r=-.22$). Trainees who scored themselves lower in this item scored higher in the examination. As was discussed in Chapter three, the

validation studies in the literature supported the proposition that those who demonstrate a higher level of competence appraise their own skills as more modest.

For year two, there was a moderate, statistically significant relationship between items in a variety of curriculum domains. These included continuity of care using a whole-person approach from *Domain one* ($r=.32$), applying principles of primary care from *Domain two* ($r=.32$), the relationship between disease patterns and general practice from *Domain three* ($r=.26$) and establishing one's own professional support networks from *Domain four* ($r=.28$). In each of these instances, trainees who scored themselves higher gained a slightly higher examination score.

During the same year, there was also a weak statistically significant relationship between self-assessed competence in particular patient groups and examination scores. Trainees who scored themselves higher in the care of adolescents ($r=.32$), older people (.26), infants ($r=.25$), indigenous people ($r=.24$) and neonates ($r=.14$), gained slightly higher average examination scores. In year three, trainees who scored themselves higher on respiratory problems, scored lower in the examination ($r=-.19$). The relationships between workload and examination results were examined and only a weak, negative correlation was found on the management interview. This meant that the trainees who saw fewer patients did better on this item.

In general, the effort to achieve consistency between self-assessment findings and results from the Fellowship Examination was always going to be limited due to the method used. Just how much correspondence would be required to achieve a modest correlation has been shown by Falchikov and Boud (1989). The only noteworthy finding on self-assessed competence and examination results was

that of undertaking a public health role. In this instance, trainees who scored themselves lower on this item, scored higher in the examination. Although the correlation was weak, it suggests that those trainees who were more competent, assessed their skills more modestly.

Satisfaction

For year one, satisfaction with remuneration received and other terms and conditions ($r=.03$) were the only items with any relationship between satisfaction findings and examination results. Those who are slightly more satisfied had slightly higher scores.

For year three, trainees who were less satisfied with video and audio consultation reviews scored higher in the examination ($r=-.22$). Those who were more satisfied with administration, scored higher on the examination ($r=.19$).

Findings in Relation to Study Hypotheses

This chapter concludes with an overview of the hypothesis in relation to the study findings. Generally confirmed were hypothesis one, that trainees' self-assessments would increase as they progressed and hypothesis six, that satisfaction would increase as trainees progressed.

Hypothesis two on gender and self-assessed competence was not confirmed for there were differences between males and females; however, the differences were small. The same applied to hypothesis seven on satisfaction and gender. Males were slightly more satisfied with administrative aspects of training, whereas females were more satisfied with opportunities to learn about population health. In terms of hypothesis three on locality and self-assessed competence, rural and urban trainees reported similar self-assessed ratings. Although there were a few

exceptions, the differences were so small that the hypothesis of little difference between urban and rural trainees was confirmed.

In relation to hypothesis eight on similarities of satisfaction ratings between urban and rural trainees, the hypothesis was not confirmed.

Hypothesis four on the decreasing of self-assessed competence ratings of trainees with certain numbers of patients was confirmed. At the point of 161 patients and over self-assessed ratings decreased. Hypothesis nine on the relationship between the number of patients seen and satisfaction ratings produced the same findings. Trainees seeing 161 patients and above, reported lower levels of satisfaction.

In relation to hypothesis five on the longitudinal cohort of trainees and the increasing of self-assessed competence ratings, the hypothesis was confirmed. Satisfaction increased in the longitudinal cohort and confirmed hypothesis ten.

A relationship was discovered between self-assessed competence and satisfaction and confirmed hypothesis eleven that anticipated that the extent to which trainees were satisfied might be associated with assessments of competence.

Hypothesis twelve was confirmed, for trainees' assessments of their own competence were shown to display little concordance with examination results.

Hypothesis thirteen on the utility of self-assessment is discussed in Chapter seven. The implications of the study findings are discussed in Chapter seven and placed into the context of the literature.

Chapter 7: Discussion of the Self-assessment Study

This chapter provides a discussion of the conceptual basis of the self-assessment study in Chapters five and six. It compares the methodology of the model in Figure 1 on page 94 and places the results of the self-assessment study in the context of the literature on the perspectives of learners and medical educators. The focus of the dissertation, the emphasis on perspectives of learners as a theoretical position, and contributions of the study feature as well as areas for further research. The conclusion considers important findings, points out some of the limitations, and identifies signposts for future direction with regard to future research. Finally, the chapter considers the significance of the study in relation to its implications for education and self-assessment practices.

Reflections

The study began with the assumption that the skills established for general practitioners represented a useful vantage point from which to examine the value of the general practice vocational training program from the perspective of learners. Chapter three analyzed the renewed and continuing interest in the perspectives of learners emphasizing two influential reviews on self-assessment on different applications of self-assessment (Gordon, 1991, 1992). These reviews included self-assessment for evaluation purposes, and initiatives to teach and validate self-assessment. In addition, Gordon highlighted some more current self-assessment frameworks that were of particular interest for their contribution to the field of educational evaluation.

Gordon's typology of self-assessment studies provided the method with which to analyse the literature. Arising from this analysis, was an argument on the benefits of self-assessment for developing professionals. At this stage, self-assessment acts as a useful starting point for a dialogue around professional lapses. The ability for trainees to evaluate what they are doing might potentially make them more secure learners, and ultimately safer practitioners.

The analysis also produced a model that expanded on the findings of Arnold et al. (1985) and provided the basis for the self-assessment study in this dissertation. In using this empirical model, the study examined aspects of general practice vocational training from the perspective of learners, and considered the role and significance of self-assessment taking into account measurable and observed aspects of the relationship between self-assessed competence and satisfaction. The overview of general practice vocational training depicted the program as based on significant and respected educational theories of adult learning. From these reviews, it was possible to draw attention to the differing purposes of self-assessment. In particular, this work contributed to new understandings of self-assessment in terms of its benefits as a collaborative educational process.

Study Research Questions and Contribution

The research questions of the self-assessment study were around aspects of general practice vocational training from the perspective of learners. The main research questions of the study examined the progression of trainees' self-assessed competence and satisfaction, and the relationship between self-assessed competence and satisfaction, in relation to stage of training and several demographic variables such as gender, locality and workload. It was the method used for the study, particularly the longitudinal survey design, that provided the study with its novel aspects. Although the method has been widely utilized in the literature on self-assessment (Arnold et al., 1985; Biro et al., 1993; Bligh

& Slade, 1996; Blumenthal et al., 2001; Clack, 1994; Donnelly et al., 1994), this particular study focused on the use of self-assessment in the context of educational evaluation. This was of particular relevance in considering the appropriateness of the program by examining the educational outcomes of a group. Over- and under-reporting might have occurred in this self-assessment study, as respondents might exaggerate their self-assessments based upon the need to leave their self-esteem intact. However, an important determining factor in choosing the self-assessment methodology was that anonymous, adult learners (without peers to compare themselves to) would not suffer socially by making a more realistic appraisal of their skills, and a critical assessment of the learning process, the curriculum and their supervisors. Of equal importance was that the longitudinal design of the study would make it acceptable to compare findings from year to year on the assumption that whatever over- and under-reporting there might be would remain consistent as has been shown in smoking prevalence data (Siahpush et al., 2002).

The central questions of the self-assessment study were whether participation in the general practice vocational training resulted in a measurable increase of self-assessed competence and satisfaction. Although this represented a seemingly straightforward task, due to the specific nature of the questions posed, the analytic techniques, coupled with the longitudinal method provided a consistent set of findings across the three years. The findings could potentially contribute to future revisions to general practice vocational training.

Another question for the self-assessment study was around the role and significance of self-assessment in medical education. The exploration of this represented a more open-ended question, considered in relation to the development of a model in Figure 1 that extended the previous methodologies used in selected and significant debates within the educational literature. The

origin of the model was demonstrated through the need for lifelong learning as shown in the historical account of general practice vocational training in Australia presented by Wilde (1998). Also viewed as essential for lifelong learning, the more recent concept of mindful practice, is predicated on the ability to reflect upon individual strengths and weaknesses (Epstein, 1999). If the primary objective of medical vocational training is the ability to engage in unsupervised general practice, it is therefore reasonable to suggest that a vocational training program that does not focus on encouraging and promoting mindful practice leaves trainees with a deficiency. Even when trainees are certified as competent, they need to be able to appraise, and reflect on their own strengths and weaknesses in order to maintain levels of competence. Herein lies the important concept that informed the rationale for the use and interpretation of self-assessment in this study. Questions around the ability to assess one's own strengths and weaknesses, and the contribution from this self-assessment study can be considered as part of the dialogue around self-assessment—especially the utility of self-assessment as identified in Chapter three, and the connection between self-assessed competence and satisfaction of trainees.

Methodology

In any study, there are multiple pathways toward answering a research question. The longitudinal research design developed and applied to the self-assessment study was appropriate for answering questions from the perspective of the learner. The model provided the pathway for a series of testable hypotheses and the expansion of the model generated a hypothesis on the role and significance of self-assessment in relation to general practice vocational training. The contextual information from the history of general practice vocational training and its evaluation clarified the rationale for the choice of methodology that was based upon feasibility of the project and the need to distinguish it from the program

evaluation work of the Outcomes Evaluation Unit. With a high response rate of 85 percent, the study was representative, and it was possible to infer from the findings to the population of current and former trainees.

The other benefit of the choice of quantitative methodology was that it facilitated the construction of a reliable scale of self-assessed competence. In asserting the value of the general practice vocational training program, it was suggested that respondents who achieved at least 50 percent competence on multiple items perceived themselves to be more competent than not. The scale therefore provided baseline data for recognizing that the program had value in the eyes of its participants. With substantial and continuing changes to the system of vocational training, the findings of this study are available for both program developers and medical educators of the current program. Both the findings and their interpretation can inform future evaluation and training. The use of longitudinal data considerably enhanced the robustness of the method. This approach offered some scope for innovation because its focus was not absolute values so necessary to the certification of practitioners as safe to practice.

Description of the study design featured an explanation of face, content and construct validity. The decision to consider these forms of validity was based on the recommendations of Feinstein (1996) who suggested that the issue be addressed simply. The focus on validity that emerged from the other literature was a counterpoint to this advice. An examination of these studies established that two measures might be claimed as valid. The measures might capture the concepts that they purport to, and be appropriately applied. This does not necessarily mean that testing a measure against another will yield high correlations. Nor does it demonstrate that two concepts measure the same

concept. Generally, the validity studies in Chapter three, and in particular the work of Feinstein (1996), established the difficulty of confirming the findings of self-assessment data with more objective sources of data. Some caution is therefore advisable, when making use of examination results to validate self-assessments.

In addition, with the time-based variables, such as stage of training, the focus of the findings was on improvements in self-ratings. All of these aspects of the self-assessment study potentially contribute to the discourse surrounding the state of knowledge with regard to educational evaluation methods.

Limitations

To benefit from a research project, and highlight areas that might advance further research, the limitations of a project should be explicit. The study in this dissertation, posed specific questions with measurable answers based upon a model extended from the literature and research findings. This was reasonable given that the approach focused on the factors that might enhance professional education.

The approach of Tamblyn (1994), that included a focus on patient care and the maintenance of clinicians' skills, was considered too broad for this doctoral project. The type of research design used for the self-assessment study was therefore selected because of a range of factors, one of which was the opportunity offered by being able to conduct a national survey of general practice trainees about their learning. The logistics of the self-assessment study of a large general practice program, in a geographically dispersed country, were such that it was only possible to triangulate data to a limited degree. The inclusion of other perspectives was evident in the examination of previous research. Some validation of self-assessment data with examination results occurred; but as the

literature strongly suggested, there was likely to be little correlation between these two forms of data. In this instance, the independently designed items in the examination were not necessarily amenable to comparison with self-assessed data. With the processes of examination question design and questionnaire design occurring very separately, they are unlikely to coincide enough for the data to display correlations. Without viewing the precise items from the bank of examination questions, it is not possible to say whether the items match. Had the examination items mirrored the questionnaire items, and had trainee self-assessments corresponded to examination data, a correlation method might still not have been able to identify concordance between the two data sources. The longitudinal aspect of the research design was an attempt to address the inability to triangulate the self-assessment data fully.

Chapter five described the research strategy, along with boundary-breaking nature of multiple method educational program evaluation designs. Also considered in Chapter five was the difficulty in educational research of bringing together disparate data sources into a coherent picture. In this instance, the missing piece of the puzzle on self-assessment came from significant observations in the literature, particularly from Chapter four on perspectives of medical educators and supervisors and the need for reflective practice. This approach offered the advantage that only the collection of trainee questionnaire data was required. The context of the project was partly a matter of timing for completion of the dissertation occurred after the system for training delivery in Australia had been devolved to multiple regionalized programs and away from the Royal Australian College of General Practitioners. Although it was not possible to consult program developers, medical educators, supervisors and particularly program directors on the findings, the Director of Education of the Royal Australian College of General Practitioners, at the time of the study offered comments on this dissertation. This supported the face validity of the findings.

During the course of the self-assessment study, no one questioned my status as a non-clinician in conducting this research. This was curious since, if one takes an insider versus outsider perspective as outlined by Becker (1993) it might have appeared more appropriate to have a person with a clinical background conducting the research. In response to this idea from Becker, the reference groups of clinically trained people provided their valuable insights and expertise to the project. It was equally important to include social science and educational methods to flesh out the range of perspectives and uncover hitherto unconsidered approaches to the data.

Perspectives of Learners and Self-assessed Competence Findings

The study with trainees of general practice vocational training sought to measure self-assessed competence in important curriculum areas, and satisfaction with aspects of the program, to ascertain its value from the perspective of learners. Some priorities influenced the reporting of the findings in order to place most emphasis on the five domains of general practice. This approach of prioritising particular aspects of practice was consistent with that of Southgate (1994) in her definitions of competence and medical practice.

Simple frequencies for self-assessed competence in five domains of general practice, and the care of particular patient groups and patient presentations provided a broad overview of the findings. In the main, trainees viewed themselves as competent. They reported themselves as being competent in *Domain one: patient-general practitioner communication*, *Domain two: applied professional knowledge and skills*, and *Domain four: professional and clinical role*. The areas in which trainees deemed themselves to be most competent (such as patient communication) reinforced the importance of evaluating the role and impact of medical education in vocational training. This is important because,

similar to nursing (Benner, 1984), considerable effort continues to be expended in this area in general practice vocational training.

Particular areas of self-assessed weakness, however, such as the management of people with drug and alcohol problems, dermatological problems, major trauma and disasters, and the care of neonates and indigenous people, are also of interest. The same applied to practice management in *Domain five: organizational and legal dimensions*, in which trainees reported the lowest self-assessed ratings of all items. These findings suggest that if course developers of general practice vocational training expended effort in these areas, their efforts to enhance trainees' confidence might be welcome.

Two items of *Domain three*: understanding the relationship between patterns and prevalence of disease and general practice, and undertaking a public health role, were areas in which respondents reported moderate weakness. The implications of these findings are particularly pertinent to the use of self-assessment.

If trainees do not feel capable in areas related to public health, for example, they will be less likely to adopt strategies that depend on general practitioners for dissemination. Evidence-based medicine and population approaches are examples of such strategies. These findings might have consequences for the implementation of public health endeavours such as campaigns encouraging patients to contemplate quitting smoking, for example. It is possible, therefore, to view self-assessment as a subjective measure, and as an indicator of general practitioner trainees' willingness to engage in particular behaviours. This builds upon the work by Cochran and Spears (1980) Donnelly et al. (1994), and Speechley et al. (1993) that presented self-assessment of competence as a proxy for willingness to practice.

The care of people with HIV/AIDS represented a different situation with the lowest self-assessed competence ratings of all items. There is, however, some room for interpretation of this finding, because this was the only area in which a sizeable number of respondents reported neither agreement nor disagreement in terms of their self-assessed competency. With this neutrality, trainees might be voicing a lack of preparedness or willingness to deal with the issue. Alternatively, people with HIV/AIDS might be infrequently presenting to trainees. Whatever interpretation prevails here suggests a need to provide exposure to HIV/patients to ensure that trainees view themselves as able to care for them.

Generally, the findings in this study were consistent with previous research, such as Donnelly et al. (1994) and Speechley et al. (1993), which highlighted weaknesses in clinical domains such as dermatology and technical skills. They also reinforced previous findings on Australian general practice vocational training in its initial evaluation (Heffernan, 1981). It is of interest that there were consistencies with the previous research—conducted almost 20 years earlier than the self-assessment study. This suggests that there will be areas in which trainees will generally require more guidance and experience even after vocational training. The implications of these findings support lifelong learning and continuing medical education.

Stage of Training

Hypothesis one of this study anticipated that as trainees progressed through training, self-assessed competence in curriculum items would increase. The data on cohorts of trainees at different stages of training supported this hypothesis. Generally, trainees in the hospital year viewed themselves as less competent in most areas of the general practice vocational training curriculum compared to those in other stages of training. Trainees' mean self-assessed ratings for the

domains of general practice, for example, increased an average of 10 percent. For the management of particular patient groups, trainees' mean ratings increased an average of 13 percent, and for the management of particular patient presentations, their mean ratings increased 12 percent.

The items relating to the provision of continuity of care using a whole-person approach (in *Domain one*), and the ordering of cost-effective investigations and managing a wide range of health and medical problems (in *Domain two*), warrant particular attention for the large statistically significant differences between hospital and other trainees. The same differences applied to understanding the relationship between patterns and prevalence of disease and general practice (in *Domain three*), establishing professional networks (in *Domain four*), and the application of relevant legal principles to practice (in *Domain five*). The implications of these findings are noteworthy, for they draw attention to the potential to enhance self-assessment skills through training in areas such as public health, peer learning and awareness of legal principles.

As established in the pilot study, the self-assessment study findings supported the use of stage of training as an independent variable with which to examine program change over time. The findings suggest the ability of adult learners to be cognizant of their strengths and weaknesses in an educational or evaluative context because they have demonstrated an ability to differentiate between them. The salient question of trainees as highlighted by Regehr et al. (1996), was not how capable or incapable trainees are in relation to their peers; but in which areas trainees' work needed the most attention in relation to their other skills. It is through this strategy that data might facilitate the offer of feedback to trainees in the form of in-training assessment to trainees about their progress as they move through different stages of training. There have been attempts to implement in training assessment in general practice vocational training (Hays & Wellard,

1998). This question from Regehr et al. (1996) could provide the starting point for an initiative to teach and use self-assessment during in training assessment. The results on stage of training in the self-assessment study suggest that adult learners might be able to reflect upon their strengths and weaknesses and that this information might be fruitfully used in this way.

Gender

In the model, gender was considered to be one of the key independent variables that might determine self-assessed competence. Hypothesis two predicted that males and females might report similar attitudes in terms of self-assessed competence in relation to items in the general practice vocational training curriculum.

In the self-assessment study, few gender differences in self-assessed competence emerged. This suggests that the findings did not support this particular hypothesis. Whereas females viewed their strengths in effectively maintaining and using patient records in years one to three, males reported their strengths as being in practice management. These differences, although not large, might suggest men identify more with a management role. Gender differences in the management of particular patient groups further supported this. Women viewed themselves as more capable in the care of neonates, infants and children. Not surprisingly, men viewed themselves as more competent in the care of men in years one and two, and women viewed themselves as more competent in the care of women. It is of note, however, that men in years two and three viewed themselves as more competent than men in the care of older people and those from rural and remote areas. If this means that women have less exposure to these problems, and, in turn, are less willing to practise in rural and remote areas, this has implications for the needs of the future rural workforce, as female trainees currently, and

increasingly, form the majority in the population of general practice trainees. This information potentially contributes to future policy debates on the issues of locality and gender.

The findings in these areas are similar to research reports of gender differences in the practice patterns, and patient populations of trainees. Britt and colleagues (2000), for example, examined gender differences in patient presentations seen by trainees, and established that women trainees saw more female-specific problems, even after adjusting for other confounding variables. These different patient presentations might account for the slightly different self-assessed competence findings for male and female trainees, simply because trainees were more confident and willing to manage patient problems and presentations in which they are more experienced and hence have less anxiety about.

The implication of these small gender differences is that, as the population of females in general practice vocational training increases there will be consequences for future training delivery. This will be particularly pertinent if women are resistant to practising in remote, rural areas, or are not receiving a mix of patient presentations to ensure an appropriate educational balance. Initiatives that inform trainees about their clinical strengths and weaknesses should be welcome in an attempt to address this aspect of training. An evaluation system that includes timely feedback would be ideal for both trainees and supervisors, and the baseline data from the self-assessment study provide signposts on the sort of data to collect to facilitate the implementation of this kind of evaluation system.

Locality

Locality was the third independent variable in the self-assessment model. Generally, the study findings confirmed the third hypothesis relating to the

similarity of the self-assessed competence ratings of urban and rural practitioners; but there were some exceptions. Rural practitioners viewed themselves as more competent in providing continuity of care using a whole-person approach, and engaging in ongoing, self-directed learning and professional development—but only in year two of the study. Rural practitioners, in year two of the study, viewed themselves as more competent in the care of adolescents, women, men, and indigenous people. In year three, they viewed themselves as more competent in the care of patients from remote, rural communities. The same results for rural trainees in year two applied to the management of sexually transmitted diseases, drug and alcohol problems and major trauma and disasters. These results on locality might relate to the different exposure of rural trainees to these patient groups with less access to specialized care. They also suggest the importance of recognising different training needs, with rural general practitioners needing to tackle a broader range of problems than their metropolitan counterparts.

In Australia, Britt, Miller and Valenti (2001) and Humphreys et al. (2003) found different patient presentations between rural general practitioners and their urban counterparts. The current documented shortfall of rural practitioners suggests a need to address this area. Over the years, the Commonwealth Department of Health and Ageing has sought to encourage the recruitment of rural general practitioners with a range of initiatives. The findings from this study can be viewed in the context of a study by Kamien (2004) on the viability of general practice. His study characterizes rural general practitioners as overworked, and in an increasingly older age group, with a larger proportion of them intending to leave general practice. Initiatives that support and encourage rural practitioners in their day-to-day work would be, therefore, an important necessity for the future health of rural communities.

Workload

The workload of trainees was the first intervening variable in the model.

Hypothesis four anticipated that self-assessed competence ratings could decrease once trainees reached a certain patient caseload. The findings of the self-assessment study confirmed this hypothesis. The findings also specifically suggested that self-assessed competence ratings decreased in the group of trainees who routinely saw more than 160 patients per week.

Trainees in the study, seeing between 120 and 160 patients, viewed themselves as highly competent in all domains except *Domain three*. The findings on workload in the self-assessment study were consistent with previous research. Fleming (1986) found a need for balance in terms of the workload of trainees. Other findings in the literature demonstrated the importance of workload to training outcomes. In a study to examine evidence of progressive learning by general practitioner trainees, Richardson, Howie and Berkeley (1974) observed that too high a patient load was not conducive to positive learning experiences. Similarly, Epstein et al. (1998) found that general practice vocational training was affected by the mix of patient population, the continuity of patients and the number and quality of interactions between supervisor and trainee.

In the self-assessment study, the findings on the importance of workload emerged as an indicator of how to achieve an appropriate balance of clinical experience and educational activity. Not only does the general practice vocational training program have particular requirements to facilitate the optimal balance between clinical and educational releases to ensure that supervisors as employers of trainees do not exploit them, but it also stipulates that students need a sufficient number of patients to enhance their skills. The study findings could be used to guide the selection of practices and supervisors, to examine what changes should

be made to the curriculum and training delivery, and the process of learning (Clack, 1994; Hojat, Borenstein, & Veloski, 1988). This is clearly an area for medical educators to continue to monitor closely, particularly in the light of the changes to training delivery since 2002 in Australia.

Longitudinal Results

Hypothesis five predicted that similar to the annual survey respondents, the self-assessed competence rating of a longitudinal panel study of trainees would increase as they progressed through the program. It offered a method to ascertain whether cross-sectional population surveys, conducted each year, might be sufficient to gain feedback on the general practice vocational training program. This would be preferable to attempting to track a group of the same individuals through the program in a panel study.

In *Domain one* and *Domain two*, findings from the longitudinal cohort of trainees broadly corresponded to the population results. In the case of *Domain three*, however, the cohort of trainees improved most in their area of greatest weakness—understanding the relationship between patterns and prevalence of disease and general practice. The same applied to the item establishing professional networks in *Domain four*, and in applying legal principles to general practice in *Domain five*. In areas of strength, the results of the cohort were more similar to those of the population. The results of the study were consistent with the work of Arnold et al. (1985) and Calhoun et al. (1990). Overall, the results suggested that the benefits of a panel study exceeded the difficulties of recruiting the same cohort over three years. However, the level of evaluation resources available would determine the decision of whether to adopt a panel study approach. If a feedback system is desirable for trainees, then a panel design is necessary. If, on the other hand, the task is simply to examine how well a program is functioning from the perspective of learners, a cross-sectional design should suffice.

Perspectives of Learners and Satisfaction Findings

In the model, satisfaction with work conditions was the second intervening variable pertained to communication factors, and work conditions. Hypothesis six predicted that satisfaction would increase as trainees progressed through training. Broadly speaking, the self-assessment study confirmed this hypothesis because trainees at later stages of training were more satisfied in all areas.

The findings were consistent with the work of Clack (1994) who advocated that satisfaction data be used to enable decisions about program changes in general practice training, delivery, and curriculum. From the perspective of learners, this demonstrates that satisfaction might act as a measure of the extent to which trainees consider a training program to be appropriate to their needs. This underscores the potential importance of satisfaction.

Hypothesis seven anticipated that males and female trainees would report similar attitudes with regard to satisfaction with training. The findings did not support this hypothesis because women reported being slightly more satisfied in areas pertaining to administrative aspects of training, and men reported being slightly more satisfied with opportunities to learn about population health. These findings suggest a relationship between satisfaction and familiarity or comfort with particular areas for they resemble research on general practice areas, patient presentations and practice patterns as identified by Britt et al. (2000).

Hypothesis eight predicted that urban and rural trainees would report similar attitudes with regard to satisfaction. Some differences emerged between rural and urban trainees: namely, that urban trainees were somewhat more satisfied with access to educational material and resources, the range of clinical experiences and the location of training posts. Therefore, the findings did not support this hypothesis. This was also borne out in previous research that has shown that rural

and remote registrars were more likely to be satisfied with the location of training posts and level of remuneration, than were their urban counterparts (Goldman et al., 2001a). On reflection, the study was unlikely to confirm hypothesis eight. Nevertheless, the expectation that there would be large differences between rural and urban trainees might be an artefact produced by the context of the development of rural training policy, particularly the formation of the Australian College of Rural and Remote Medicine (Hutton-Czapski, 2000).

Hypothesis nine predicted that satisfaction ratings would decrease once trainees reached a specific patient caseload. The results on the relationship between workload and satisfaction suggested dissatisfaction on particular items such as the range of clinical experience among trainees with a caseload of over 121, and confirmed the hypothesis. This finding offers an important method for program managers and designers wishing to monitor their programs.

The findings on satisfaction also confirmed hypothesis 10 on the longitudinal cohort of trainees. Most important, there was a consistent pattern of improved satisfaction as the same trainees progressed through the program. Satisfaction improved most in areas where trainees previously reported least satisfaction. Notably, trainees in the longitudinal cohort demonstrated larger improvements in their satisfaction ratings—improving most in opportunities to learn about population health in the context of general practice. These findings were similar to those of Arnold et al. (1985); however, the majority of other research identified was cross-sectional in its design. The item of satisfaction with opportunities to learn about public health reflected an increased awareness in this area as trainees progressed. As already noted, the significance of this finding has lessons for those developing public health programs aimed at General Practitioners in terms of their preferences and willingness to implement public health measures.

Hypothesis 11 predicted that satisfaction ratings might determine self-assessed competence. A regression model incorporating additive scales measured this. The scales consisted of opportunities to learn communication skills, to learn about public health in the context of general practice, to meet with other trainees, to engage in external clinical teaching sessions, and video and consultation reviews as items belonging to a scale encompassing satisfaction with communication. This scale featured in the multiple regression analysis and the results identified that satisfaction with communication factors was the strongest predictor of self-assessed competence, particularly in year three.

Although it might be tempting to de-emphasize satisfaction as a desirable but not a necessary prerequisite for effective learning, these findings on stage of training, locality, workload and particularly the relationship between satisfaction and self-assessed competence suggested that satisfaction was a focal point of the self-assessment study. If satisfaction is as important as these data suggest, this opens the way for interventions that might facilitate improved learning outcomes by addressing the morale of learners as they progress. Particular benefit might be gained from examining the perspective of learners by closely monitoring the number of patients seen by trainees: if too many patients are seen, trainees do not have a chance to benefit fully from their educational experiences outside their general practice placement; if too few patients are seen, on the other hand, trainees will not gain access to a broad range of clinical experience. In addition, it might be valuable to address opportunities to learn communication skills and population health as special items that feature extensively in the data. These items might be amenable to change and therefore influence trainee assessments.

Overall, the study established that participation in general practice vocational training accompanied an increase of self-assessed competence in areas of the curriculum, and that trainees were generally satisfied with their training

experiences. Whereas the possible increasing maturity of trainees should be considered as an interpretation of the findings, it is a less likely consideration with adult learners. The literature on sub-group differences suggested that examining group differences is an accepted method for evaluating the appropriateness of training. This is particularly the case in a situation such in Australia, in which training has become the major pathway to vocational registration for General Practitioners, and control groups for randomised controlled trials no longer exist. For these reasons, it is reasonable to some extent to attribute the findings to the program. As well as answering the research questions on self-assessed competence and satisfaction, the questions on stage of training, gender, locality and workload provided insights into the utility of trainee self-assessments, particularly in relation to their strengths and weaknesses.

A brief history of the Australian general practice vocational training program demonstrated the importance that founders placed on the role of self-assessment and self-directed learning. In light of this, it was of particular note that the magnitude of increase in self-assessed competence, as trainees progressed through their training, was high in areas of special weakness, such as the management of people with dermatological problems, and in public health items such as understanding the relationship between patterns and prevalence of disease and general practice. These findings confirmed the importance of stage of training—the second independent variable in the model. Overall, these findings also confirmed the usefulness of having a model for an area such as educational evaluation particularly in as contested an area as general practice vocational training. Without over-interpreting these data, they do suggest that the general practice vocational training program successfully met its objective of enhanced confidence. This objective has been a central part of the program from the start, and is consistent with its humanist orientation. As a more recent concept to the arena of medical education, Epstein (1999) defined practitioners' self-assessment

of their own strengths and weaknesses as important for mindful practice. Although Epstein related the benefits of mindful practice to a practitioner's ability to exercise empathy and compassion, its benefits could be measurable across a range of competencies. By facilitating increasing confidence as trainees progressed, the general practice vocational training program contributes to habits of lifelong learning, one of the stated goals of mindful practice. According to Epstein, the humanist values required to exercise empathy and compassion are relevant to any aspect of clinical care. While an educational model that facilitates mindful practice in trainees is usually one that entails an appraisal of an individual, it was not part of this study (and the overall research orientation) to communicate the findings back to the individual general practice trainee. Nevertheless, this could quite feasibly occur as part of the teaching context, whereby the skills and habits of self-assessment were established and developed.

Perspectives of Educators and Study Findings

The depiction of the perspectives of medical educators as well as trainees in the model in Figure 1 supports the conceptual basis of the thesis on the utility of self-assessment. The consideration of reflective learning and adult learning in relation to observable aspects of the model and the quantitative findings of the self-assessment study affirmed the approach as a novel one for seeking to include multiple perspectives in considering the value of self-assessment. It also supported Hypothesis 13 on the usefulness of self-assessment as an educational tool.

The process of incorporating the perspective of educators began with the expression of particular theoretical positions around the value of self-assessment in education. The purpose of this was to augment the findings of the self-assessment study. The beginning of Chapter four explored whether knowledge derived from the subjectivity of experience, as well as reason. This represented

a key plank of educational philosophy in the 1960s and the 1970s and was a pathway to the recognition of the importance of experiential learning. The importance of the examination of this history was to provide some understanding of the place of self-assessment in general practice vocational training. Chapter four also related the origins of experiential learning to the need to concede mistakes as part of the learning process in medical disciplines, while recognizing the need to preserve the role of the medical practitioner as healer. The findings from this self-assessment study on self-assessed competence and variables such as stage of training serve to contribute to this body of literature that supports the enhancement of self-assessment skills in trainees.

Chapter four as an extension of the model highlighted the influence of Schön on the development of general practice vocational training in Australia, in his emphasis on developing the reflective practitioner as a significant part of learning. The concept of reflection-in-action, as depicted by medical educators in describing learning in the practice, as well as the role of coaching by supervisors was also emphasised in the literature (Gruppen et al., 2000). The relationship of these two concepts presented reflection-in-action as a reframing process, facilitated by a general practice supervisor who acts as a critical friend. In this context, several aspects emerged from the findings of the self-assessment study with regard to learners' reactions to the form of general practice training used in Australia. The reasonably high mean scores on particular satisfaction items such as learning challenges provided, role models provided by supervisors and supervision and advice received supported the research on the importance of the supervisor in the learning interaction. They suggest that self-assessment does not occur in isolation; but requires the help of a critical friend. The findings on the relationship between self-assessment ratings and satisfaction further underscore the importance of the role of satisfaction in the learning interaction.

Reading Schön's reflection-in-action as a theory of meta-cognition, emphasizes the centrality and importance of observation and feedback in general practice training. A clear way of explaining this, is with the concept of knowing about knowing. Also termed meta-cognition, many general practice vocational training curricula have articulated this general concept. Terminology that is even more current is that of mindful practice. This is when practitioners are conscious of their own thinking and feeling process while engaged in practice. Education and training facilitate this fully conscious state of mind. The activity continues in the consulting room by helping the professional to ask the following questions: 'How do I respond to uncertainty? How do I evaluate risks? How do I self-correct when doing a difficult procedure? (Epstein, 1999). These questions provide some guidance and lessen the ambiguity around reflection-in-action. They also build upon additional work from Epstein and Hundert (2002) that placed self-assessment in a category of habits of mind necessary for family practice.

It is possible to see the connection between the concept of mindful practice and Schön's concept of reflective practice. The design of initiatives for the teaching of self-assessment might readily apply the findings of the self-assessment study. This is because the trainees in the study were not engaged in a long intervention for the teaching of self-assessment. The three questions from Epstein (1999) are a good starting point for training reflective practitioners.

The findings of the self-assessment study also relate to many of the principles of Knowles (1984). These principles contribute to the observable aspects of the model upon which the self-assessment study was based. Whereas idealism was the foundation of general practice vocational training in Australia, the initial altruism of the program diminished due to a wide range of influences. The imperative of the summative examination supplanted Knowles' principle of the need to know and its emphasis on general practice trainees' curiosity of,

and commitment to, the discipline of medicine. Nevertheless, an implication of these study findings around differences in self-assessed ratings is that program designers could further develop self-assessment, particularly as a way of resolving the tension that might occur between supervisor and trainee on the matter of workloads. Trainees could further develop their self-assessment skills. Supervisors, too, could use self-assessment to monitor their teaching and supervision. This recommendation represents a contribution that arose from the use of the model upon which the self-assessment study was based.

Another principle from Knowles (1984) that was considered was the importance of the recognition of learners' experiences. In the qualitative research examined in Chapter four, satisfaction with supervision was relatively high. However, questions arose as to the capability of supervisors to provide sufficient appropriate clinical experience and act as effective educators, when they have little training in the area. This was evident in the trainees' perceptions of the rather limited repertoire of teaching activities that supervisors engaged in. Further research in the area would clarify the extent of this issue, and provide ways of addressing it. Although the self-assessment study occurred under the system of training of the Royal Australian College of General Practitioners, subsequent studies could address the issue of supervision and compare current data with that from this self-assessment study.

Changes to the general practice vocational training program over time have not diminished the importance of the ethos of valuing the learner's experience; however, other issues have attained equal importance. Factors pertaining to workforce issues such as rural training posts, and terms and conditions as outlined in the Introduction, particularly with regard to limitations on the number of training places, have influenced the numbers of trainees. These items attracted relatively modest satisfaction ratings in the self-assessment study. The model

as the basis of the self-assessment study represented a way of addressing these perspectives. The next step for such research would be to conduct in depth research of the perspectives of supervisors on the issues of trainee satisfaction.

The derivation of lifelong learning in the context of general practice training featured in the history of general practice vocational training. This section posed the question of why self-assessment skills are important for adult learners. The model addressed supervisors' perspectives by considering literature and educational theories. The model was therefore an extension of empirical research and considered quantitative and qualitative research findings in the form of observed and measured relationships depicting self-assessment and the recognition of one's limitations as a professional. The contribution made by presenting the perspectives of educators was significant in that it augmented those of trainees by locating theories that support the use of self-assessment. In addition, the findings provide signposts on further research, particularly on the role of satisfaction.

Conclusion

The self-assessment study provided an understanding of which aspects of training, the curriculum, and the training experience to enhance. Most important, general practice trainees reported high levels of satisfaction with the range of clinical experience, learning challenges provided, the role models provided by their supervisors, and the advice received (from supervisors, mentors and other more experienced practitioners).

Although respondents in this study generally reported favourably about the program, they offered some critique of its educational methods. This was consistent with earlier data on the frequency of educational activities that

suggested that corridor or ad hoc teaching was the most frequent teaching activity of supervisors. Although there were undoubtedly outstanding teachers among supervisors of the program, the self-assessment study suggests that program managers should also work towards providing a supportive critical community around supervisors. This would serve to maintain supervisors' teaching skills, and more fully recognize their contributions. At the same time, the findings of the self-assessment study supported the value of asking trainees about their satisfaction with the program, to identify potential deficiencies to encourage the beginnings of a dialogue around establishing a supportive, critical community. As a key purpose of this study, considering trainees' perspectives continues as an important educational strategy to benefit trainees and the program as a whole. This was one of the key recommendations of the study around the role and significance of self-assessment.

The research questions generated a series of useful and worthwhile answers. Qualitative analysis of the literature highlighted the value of self-assessment as preparation for the developing professional. An important point to emerge from this exercise was the value of trainees evaluating, with some accuracy, their own capabilities as a means to making them more secure learners, and potentially safer practitioners.

There was a special focus on the benefits of measuring confidence in public health as a subjective measure, important in the design of programs. Of particular note were the measurable differences between trainees as they progress through training, preferences of males and females, and urban and rural differences, that might be of importance once training is complete, particularly in the area of lifelong learning. Findings oscillated around the need for multiple perspectives in assessing the role and value of self-assessment and general practice training. These results produced additional questions for future research to further examine

the general practice vocational training environment. Since the collection and interpretation of the data from this self-assessment study, the environment has undergone a series of changes, and will continue to evolve.

References

- Abrahamson, S. (1987). *Report of the review of the Family Medicine Programme*. Sydney: Royal Australian College of General Practitioners.
- Agyris, C., & Schön, D. (1978). *Organizational learning: A theory of action perspective*. Massachusetts: Addison-Wesley Publishing Company.
- Allen, J., & Rashid, A. (1998). What determines competence within a general practice consultation? Assessment of consultation skills using simulated surgeries. *British Journal of General Practice*, 48(5), 1259–1262.
- Andersen, N. A., Bridges-Webb, C., & Chancellor, A. (1986). *General practice in Australia* (1st ed.). Sydney: Sydney University Press.
- Antonelli, M. A. S. (1997). Accuracy of second-year medical students' self-assessment of clinical skills. *Academic Medicine*, 72(10), S63–65.
- Arnold, L., Willoughby, L. T., & Calkins, V. E. (1984). Understanding the clinical performance of physicians: a factor analysis approach. *Journal of Medical Education*, 59(7), 591–594.
- Arnold, L., Willoughby, T. L., & Calkins, V. E. (1985). Self-evaluation in undergraduate medical education: A longitudinal perspective. *Journal of Medical Education*, 60(1), 21–28.
- Australian Medical Workforce Advisory Committee. (1996). *Australian Medical Workforce Advisory Committee Annual Report*. Sydney: Australian Medical Workforce Advisory Committee.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122–137.
- Bashook, P. G., & Parboosingh, J. (1998). Recertification and the maintenance of competence. *British Medical Journal*, 316(7130), 545–548.
- Bazeley, P., & Richards, L. (2000). *The NVivo qualitative project book*. London: Sage Publications.
- Becker, H. (1993). How I learnt what a crock was. *Journal of Contemporary Ethnography*, 22(1), 28–35.
- Benner, P. (1984). *From novice to expert: Excellence and power in clinical nursing practice*. California: Addison-Wesley Publishing Company.
- Biggs, J. B., & Moore, P. J. (1993). *The process of learning*. Prentice-Hall.
- Biro, F. M., Siegel, D. M., Parker, R. M., & Gillman, M. W. (1993). A comparison of self-perceived clinical competencies in primary care residency graduates. *Pediatric Research*, 34(5), 555–559.
- Bligh, J. (1992). Independent learning among general practice trainees: an initial survey. *Medical Education*, 26(6), 497–502.

- Bligh, J., & Slade, P. (1996). A questionnaire examining learning in general practice. *Medical Education*, 30(1), 65–70.
- Blumenfeld, D. (1996). Rationalism. In J. J. Chambliss (Ed.), *Philosophy of education: An encyclopedia* (Vol. 1671, pp. 720). New York: Garland Publishing.
- Blumenthal, D., Gokhale, M., Campbell, E., & Weissman, J. (2001). Preparedness for clinical practice: Reports of graduating residents at academic health centers. *Journal of the American Medical Association*, 286(9), 1027–1033.
- Boendermaker, P. M., Schuling, J., Meyboom-de Jong, B., Zwierstra, R. P., & Metz, J. C. M. (2000). What are the characteristics of the competent general practitioner trainer? *Family Practice*, 17(6), 547–553.
- Bollen, M., & Saltman, D. (2000). *General practice in Australia: 2000*. Canberra: Department of Health and Aged Care.
- Bradley, F., Murphy, A. W., & Lambe, S. (1996). A national survey of career pathways of all graduates of vocational training programmes in GP in Ireland. *European Journal of General Practice*, 2, 157–161.
- Bridges-Webb, C. (1976). The Australian general practice morbidity and prescribing survey, 1969–1974. *Medical Journal of Australia*, 2, S1–28.
- Bridges-Webb, C., Britt, H., Miles, D., Neary, S., Charles, J., & Traynor, V. (1993). Morbidity and treatment in general practice in Australia. *Australian Family Physician*, 3(22), 336–339, 342–346.
- Britt, H., Bhasale, A., Miles, D., Meza, A., Sayer, G., & Angelis, M. (2000). The sex of the general practitioner: A comparison of characteristics, patients and medical conditions managed. *Medical Care*, 34(5), 403–415.
- Britt, H., Miller, G., Sayer, G., & Sedgwick, D. (1999). *Comparison of practice patterns using Health Insurance Commission data for GPs who have completed the RACGP Training Program and those who have not*. Melbourne: Royal Australian College of General Practitioners Outcomes Evaluation Unit.
- Britt, H., Miller, G., & Valenti, L. (2001). *It's different in the bush: A comparison of general practice activity on metropolitan and rural areas of Australia 1998–2000*. Canberra: Australian Institute of Health and Welfare.
- Brookfield, S. (1986). *Understanding and facilitating adult learning*. San Francisco: Jossey-Bass.
- Brookfield, S. (1995). *Becoming a critically reflective teacher* (1st ed.). San Francisco: Jossey-Bass.

- Brownlea, A., & Greco, M. (1998). *GPEP final report: utilising patient feedback in a general practice training program*. Griffith University and Royal Australian College of General Practitioners.
- Calhoun, J. G., Ten Haken, J., & Wooliscroft, J. O. (1990). Medical students' development and self- and peer-assessment skills: A longitudinal study. *Teaching and Learning in Medicine*, 2(1), 25–29.
- Calhoun, J. G., Wooliscroft, J. O., Ten Haken, J., Love, S. J., & Boxer, G. J. (1986). Using videotape to evaluate medical students' physical examination skills. *Medical Teacher*, 8(4), 367–372.
- Campbell, C. M., Slotnick, H. B., & Parboosingh, J. (1999). Outcomes related to physicians' practice-based learning. *The Journal of Continuing Medical Education in the Health Professions*, 19(4), 234–241.
- Campbell, L. M., & Murray, T. S. (1996). Assessment of competence. *British Journal of General Practice*, 46(411), 619–622.
- Candy, P. (1991). *Self-direction for lifelong learning: A comprehensive guide to theory and practice*. San Francisco: Jossey-Bass.
- Carlisle, C. (2000). Reflecting on confidence and competence in skills acquisition. *Medical Education*, 34(11), 886–887.
- Chambers, R., Wall, D., & Campbell, I. (1996). Stresses, coping mechanisms and job satisfaction in general practitioner registrars. *British Journal of General Practice*, 46, 343–348.
- Chien, A., & Fennessy, S. (2004). Action research as a learning tool in general practice. *Australian Family Physician*, 33(1-2), 60–61.
- Clack, G. B. (1994). Medical graduates evaluate the effectiveness of their education. *Medical Education*, 28(5), 418–431.
- Clack, G. B., & Head, J. O. (1999). Gender differences in medical graduates' assessment of their personal attributes. *Medical Education*, 33(2), 101–105.
- Cochran, S., & Spears, M. (1980). Student self-assessment and instructors' ratings: a comparison. *Journal of the American Dietetic Association*, 76(3), 253–257.
- Collard, S., & Law, M. (1989). The limits of perspective transformation: A critique of Mezirow's theory. *Adult Education Quarterly*, 39(2), 99–107.
- Commonwealth Department of Health and Aged Care (Ed.). (2000). *General practice in Australia: 2000* (1st ed.). Canberra: Australian Government Publishing Service.

- Commonwealth Department of Health and Family Services. (1996). *General Practice in Australia: 1996*. Canberra: Australian Government Publishing Service.
- Commonwealth Department of Health and Family Services. (1998a). *General practice education: The way forward. Report of the ministerial review of general practice training* (No. 2290). Canberra: Commonwealth Department of Health and Family Services.
- Commonwealth Department of Health and Family Services. (1998b). *General practice in Australia: Changing the future through partnerships. Report of the General Practice Strategy Review Group*. Canberra: Commonwealth Department of Health and Family Services.
- Cooper, B., Goldman, S., Jasper, A., & Wellard, R. (2001a). *An analysis of registrar practice and educational activities*. Melbourne: Royal Australian College of General Practitioners.
- Cooper, B., Goldman, S., Jasper, A., & Wellard, R. (2001b). *Deconstructing feedback* Paper presented at the National Medical Educator Conference, Townsville.
- Cooper, B., Jasper, A., Goldman, S., & Wellard, R. (2001c). *A study of registrar satisfaction with general practice training placements and educational activities*. Melbourne: Royal Australian College of General Practitioners.
- de Vaus, D. (2002). *Surveys in social research* (5th ed.). Crow's Nest: Allen & Unwin.
- Dean, A. (1996). Using a microcomputer for field investigation. In M. B. Gregg (Ed.), *Field Epidemiology* (pp. 217–235). New York: Oxford University Press.
- Denzin, N. (1989). *The research act: A theoretical introduction to sociological methods* (3rd ed.). New York: Prentice Hall.
- Docker, J., Carson, N., Hughes, P., Kamien, M., McGaw, B., & Nowotny, R. (1993). *Report of the knowledge, skills and attitudes task force: The Royal Australian College of General Practitioners: Family Medicine Program*.
- Donnelly, J. F., Boltri, J. M., Weeks, M. E., & Mengel, M. B. (1994). Measuring self-reported comfort with general family practice skills during a required third-year family practice clerkship. *Family Medicine*, 26(1), 21–26.
- Dowton, B. S. (2005). Postgraduate medical education: Rethinking and integrating a complex landscape. *Medical Journal of Australia*, 182(4), 177–180.
- Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. Oxford: Basil Blackwell.

- Ende, J. (1983). Feedback in clinical medical education. *Journal of the American Medical Association*, 250(6), 777–781.
- Epstein, R. M. (1999). Mindful practice. *Journal of the American Medical Association*, 282, 836–839.
- Epstein, R. M., Cole, D. R., Gawinski, B. A., Piotrowski-Lee, S., & Ruddy, N. B. (1998). How students learn from community-based preceptors. *Archives of Family Medicine*, 7(2), 149–154.
- Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing competence. *Journal of the American Medical Association*, 287(2), 226–235.
- Eraut, M. (1994). *Developing professional knowledge and competence* (1st ed.). London: The Falmer Press.
- Eraut, M. (1995). Schön shock: A case for reframing reflection-in-action. *Teachers and Teaching: Theory and Practice*, 1(1), 9–22.
- Fabb, W. E., Heffernan, M., Phillips, W., & Stone, P. (1976). *Focus on learning in family practice*. Melbourne: Royal Australian College of General Practitioners.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59(4), 395–430.
- Farnill, D., Hayes, S., & Todisco, J. (1997). Interviewing skills: Self-evaluation by medical students. *Medical Education*, 31, 122–127.
- Feinstein, A. (1996). Mathematical models and clinical practice. In J. Daly (Ed.), *Ethical intersections: Health research, methods and researcher responsibility* (pp. 203–210). St Leonards: Allen & Unwin.
- Fielding, D. W., Page, G. G., & Schulzer, M. (1982). Assuring continuing competence: identification and validation of a practice based assessment blueprint. *American Journal of Pharmaceutical Education*, 56, 21–29.
- Fincher, R.-M., Lewis, L. A., & Kuske, T. T. (1993). Relationships of interns' performances to their self-assessments of their preparedness for internship and to their academic performances in medical school. *Academic Medicine*, 68(2), S47–50.
- Fleming, D. M. (1986). A comparison of the practice activities of trainees and principals. *Journal of the Royal College of General Practitioners*, 36(286), 212–216.
- Fleming, P. (1993). *An integrated program of vocational training for general practice in Australia* (2nd ed.): Royal Australian College of General Practitioners.
- Foucault, M. (1979). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans.). Harmondsworth: Penguin.

- Fox, R. (1957). Training for uncertainty. In R. Merton, G. Reader and P. Kender (Eds.), *The student-physician* (pp. 207–241). Massachusetts: Harvard University Press.
- Fox, R. C. (1980). The evolution of medical uncertainty. *Health and Society*, 58(1), 1–49.
- Fraser, R. C., McKinley, R. K., & Mulholland, H. (1994). Consultation competence in general practice: testing the reliability of the Leicester assessment package. *British Journal of General Practice*, 44(384), 293–296.
- Freidson, E. (1970). *Profession of medicine: A study of applied knowledge*. Chicago: University of Chicago Press.
- Garrison, D. (1997). Self-directed learning: toward a comprehensive model. *Adult Education Quarterly*, 48(1), 18–33.
- Ginsburg, S., Regehr, G., Hatala, R., McNaughton, N., Frohna, A., Hodges, B., et al. (2000). Context, conflict, and resolution: A new conceptual framework for evaluating professionalism. *Academic Medicine*, 75(Suppl. 10), S6–S11.
- Girgis, A., Sanson-Fisher, R. W., & Walsh, R. (2001). Preventive medicine and other interactional skills of General Practitioners, surgeons and physicians: perceived competence and endorsement of postgraduate training. *Preventive Medicine*, 32, 78–81.
- Glasgow, R., Ory, M., Klesges, L., Cifuentes, M., Fernald, D., & Green, L. (2005). Practical and relevant self-report measures of patient health behaviors for primary care research. *Annals of Family Medicine*, 3(1), 73–81.
- Goedhuys, J., & Leroy, D. (1990). Belgian general practitioners on their vocational training: A postal inquiry. *Scandinavian Journal of Primary Health Care*, 8(4), 249–253.
- Goldman, S. (1998). National medical educator and supervisor focus group unpublished data. Unpublished Transcript.
- Goldman, S. (1999). *A study of registrar views to evaluate outcomes of the RACGP Training Program*. Melbourne: Royal Australian College of General Practitioners.
- Goldman, S. (2000a). *A pilot study to evaluate supervisor views of general practice vocational training*. Melbourne: Royal Australian College of General Practitioners.
- Goldman, S. (2000b). *Supervisor views on aspects of the RACGP Training Program*. Melbourne.

- Goldman, S., Cooper, B., Jasper, A., & Wellard, R. (2001a). *The effects of gender and locality on registrar views of the Training Program*. Melbourne: Royal Australian College of General Practitioners.
- Goldman, S., Jasper, A., & Wellard, R. (1998). *A pilot study to evaluate registrar views on outcomes of the RACGP Training Program*. Melbourne: Royal Australian College of General Practitioners.
- Goldman, S., Jasper, A., & Wellard, R. (2001b). *A further study of registrar views to evaluate aspects of general practice vocational training*. Melbourne: Royal Australian College of General Practitioners.
- Goldman, S. P., Swerissen, H., Jasper, A. A., & Wellard, R. F. (1999). Self-reporting of competence by general practitioners with vocational training. *Focus on Health Professional Education: A Multidisciplinary Journal*, 1(2), 50–59.
- Gordon, M. J. (1991). A review of the validity and accuracy of self-assessment in health professions training. *Academic Medicine*, 66(12), 763–769.
- Gordon, M. J. (1992). Self-assessment programs and their implications for health professions training. *Academic Medicine*, 67(10), 672–679.
- Gordon, M. J. (1997). Cutting the Gordonian knot: A two-part approach to the evaluation and professional development of residents. *Academic Medicine*, 72(10), 876–880.
- Graetz, B., & McAllister, I. (1988). *Dimensions of Australian society* (First ed.). South Melbourne: The MacMillan Company of Australia Pty Ltd.
- Grol, R., Tielens, V., & Mekkink, H. (1985). Attitude changes in the vocational training of general practitioners. *Medical Education*, 10, 479–486.
- Gruppen, L., White, C. B., Fitzgerald, J. T., Grum, C. M., & Wooliscroft, J. O. (2000). Medical students' self-assessments and their allocations of learning time. *Academic Medicine*, 75(4), 374–379.
- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, 43–62.
- Haug, M. (1973). An alternative hypothesis for the future. In P. Halmos (Ed.), *Professionalism and social change* (Vol. 20, pp. 338). Staffordshire: University of Keele.
- Hays, R., & Piterman, L. (2000). Education and training for General Practitioners. In Commonwealth Department of Health and Aged Care (Ed.), *General practice in Australia: 2000* (pp. 346–382). Canberra: Australian Government Publishing Service.
- Hays, R., & Wellard, R. (1998). In training assessment in postgraduate training for general practice. *Medical Education*, 32(5), 507–513.

- Hays, R., Worley, P., & Coote, B. (2004). Education and training. In *General practice in Australia: 2004* (pp. 268–301). Canberra: Australian Government Publishing Service.
- Hays, R. B., van der Vleuten, C., Fabb, W. E., & Spike, N. A. (1995). Longitudinal reliability of the Royal Australian College of General Practitioners Certification Examination. *Medical Education*, 29(4), 317–321.
- Heffernan, T. (1981). *The formative evaluation of the Family Medicine Programme Summary report one: Programwide results and recommendations from the first round of measurement*. Melbourne: The Royal Australian College of General Practitioners.
- Henbest, R. J., & Fehresen, S. G. (1985). Preliminary study at the Medical University of South Africa on student self-assessment as a means of evaluation. *Journal of Medical Education*, 60(1), 66–67.
- Hojat, M., Borenstein, B. D., & Veloski, J. J. (1988). Cognitive and noncognitive factors in predicting the clinical performance of medical school graduates. *Journal of Medical Education*, 63(4), 323–325.
- Hojat, M., Gonnella, J. S., Veloski, J. J., & Erdmann, J. B. (1996). Jefferson medical college longitudinal study: A prototype for evaluation of changes. *Education for Health*, 9(1), 99–113.
- Hojat, M., Veloski, J. J., & Borenstein, B. D. (1986). Components of clinical competence ratings of physicians: an empirical approach. *Educational and Psychological Measurement*, 46(3), 761–769.
- Hojat, M., Vogel, W. H., Zeleznik, C., & Borenstein, B. D. (1988). Effects of academic and psychosocial predictors of performance in medical school on coefficients of determination. *Psychological Reports*, 63, 383–394.
- Humphreys, J. S., Jones, J. A., Jones, M. P., Mildenhall, D., Mara, P. R., Chater, B., et al. (2003). The influence of geographical location on the complexity of rural general practice activities. *Medical Journal of Australia*, 179(8), 416–420.
- Hurley, T., & Cummins, C. (1982). *The Family Medicine Programme—report of a review*. Melbourne: Royal Australian College of General Practitioners.
- Hutton-Czapski, P. (2000). Letter from Australia. *Canadian Journal of Rural Medecine*, 5(1), 28–29.
- Jansen, J. J. M., Grol, R. P. T. M., Crebolder, H. F. J. M., Rethans, J.-J., & van der Vleuten, C. P. M. (1998). Failure of feedback to enhance self-assessment skills of general practitioners. *Teaching and Learning in Medicine*, 10(3), 145–151.

- Jansen, J. J. M., Tan, L. H. C., van der Vleuten, C. P. M., van Luijk, S. J., Rethans, J. J., & Grol, R. P. T. M. (1995). Assessment of competence in technical clinical skills of general practitioners. *Medical Education*, 29, 247–253.
- Johnson, N., Hasler, J., Mant, D., Randall, T., Jones, L., & Yudkin, P. (1993). General practice careers: changing experience of men and women vocational trainees between 1974 and 1989. *British Journal of General Practice*, 43, 141–145.
- Johnson, N. (1993). Feasibility of developing and selecting criteria for the assessment of clinical performance. *British Journal of General Practice*, 43, 599–502.
- Kamien, M. (2004). The viability of general practice in rural Australia. *Medical Journal of Australia*, 180(7), 318–319.
- Kant, I. (1781). *Critique of pure reason* (J. Meiklejohn, Trans.). London: Dent and Sons.
- Keck, J. W., & Arnold, L. (1979). Efficacy of cognitive/noncognitive measures in predicting resident-physician performance. *Journal of Medical Education*, 54(October), 759–765.
- Kefalas, B., Usherwood, T., & Knight, A. (2005). External clinical teaching visits in the regionalised environment. *Australian Family Physician*, 34(12), 10–12.
- Kegel-Flom, P. (1975). Predicting supervisor, peer, and self-ratings of intern performance. *Journal of Medical Education*, 50(8), 812–815.
- Kiley, M., Peterson, R., Mullins, G., & Wellard, R. (2000). Development and implementation of a national curriculum for postgraduate general practice training. *Education for General Practice*, 11, 268–275.
- Knowles, M. S. (1984). *The adult learner: A neglected species* (3rd. ed.). Houston: Gulf Publishing Company.
- Knowles, M. S., Holton, E. F., & Swanson, R. A. (1998). *The adult learner: The definitive classic on adult education and human resource development* (5th ed.). Massachusetts: Butterworth-Heinemann.
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. New Jersey: Prentice Hall
- Kolm, P., & Verhulst, S. J. (1987). Comparing self- and supervisor evaluations: a different view. *Evaluation and Health Professions*, 10(1), 80–89.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.

- Ley, R. (1991). *Communicating with patients* (2nd. ed. Vol. 4). Cambridge: Chapman and Hall.
- Lichtenstein, B. M. (2000). Generative knowledge and self-organized learning. *Journal of Management Inquiry*, 9(1).
- Liddell, M. J. (1999). *Communication within the consultation: Attitudes and performance of medical undergraduates at Monash University*. Monash University, Melbourne.
- Linn, B. S., Arostegui, M., & Zeppa, R. (1975). Performance rating scale for peer and self assessment. *British Journal of Medical Education*, 9(2), 98–101.
- Malinen, A. (2000). *Towards the essence of adult experiential learning*. Jyväskylä: So Phi.
- Mamede, S., & Schmidt, H. G. (2004). The structure of reflective practice in medicine. *Medical Education*, 38, 1302–1308.
- Marienau, C. (1999). Self-assessment at work: outcomes of adult learners' reflection on practice. *Adult Education Quarterly*, 49(3), 135–146.
- Martin, D., Regehr, G., Hodges, B., & McNaughton, N. (1998). Using videotaped benchmarks to improve the self-assessment ability of family practice residents. *Academic Medicine*, 73(11), 1201–1206.
- Maudsley, G., & Strivens, J. (2000a). Promoting professional knowledge, experiential learning and critical thinking for medical students. *Medical Education*, 34, 535–544.
- Maudsley, G., & Strivens, J. (2000b). 'Science', 'critical thinking' and 'competence' for tomorrow's doctors. a review of terms and concepts. *Medical Education*, 34(1), 53–60.
- McClennan, B. L., & Herlihy, C. S. (1995). The continuing competence needs of physicians: A survey of the medical specialty societies. *American Journal of Roentgenology*, 165(4), 789–795.
- Mezirow, J. (1998). On critical reflection. *Adult Education Quarterly*, 48(3), 185–198.
- Mezirow, J., & Associates. (1990). *Fostering critical reflection in adulthood: A guide to transformative and emancipatory learning* (1st ed.). San Francisco: Jossey-Bass Publishers.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65(Suppl. 9), S63–S67.
- Minsky, M. (1977). Frame system theory. In P. Johnson-Laird and P. Watson (Eds.), *Thinking: Readings in cognitive science* (pp. 355–376). New York: Cambridge University Press.

- Mitchell, M., Srinivasan, M., West, D., Franks, P., Keenan, C., Henderson, M., et al. (2005). Factors affecting physician performance: Development of a theoretical model and a focused literature review. *Academic Medicine*, 80(4), 376–389.
- Morton, J. B., & Macbeth, W. A. A. G. (1977). Correlations between staff, peer and self assessments of fourth-year students in surgery. *Medical Education*, 11(3), 167–170.
- Newble, D., Dawson, B., Dauphinee, D., Page, G., Macdonald, M., Swanson, D., et al. (1994). Guidelines for assessing clinical competence. *Teaching and Learning in Medicine*, 6(3), 213–220.
- Norman, G. (2002). Research in medical education: three decades of progress. *British Medical Journal*, 324(7353), 1560–1562.
- Norusis, P. (2002). *SPSS 11.0 Guide to data analysis*. Upper Saddle River NJ: Prentice-Hall.
- Parsons, T. (1939). The professions and social structure. *Social Forces*, 17, 457–467.
- Patton, M. Q. (1986). *Utilization-focused evaluation* (2nd. ed ed.). California: Sage Publications Inc.
- Prideaux, D. (2002). Researching the outcomes of educational interventions: A matter of design. *British Medical Journal*, 324(7330), 126–127.
- Prideaux, D., Henry-Edwards, S., Middleton, P., & Lyons-Reid, A. (1999). *Outcomes evaluation in general practice vocational training programs: A literature review*. Melbourne: Royal Australian College of General Practitioners Outcomes Evaluation Unit
- Prideaux, D., Lyons-Reid, F., & Ashenden, R. (1995). *An evaluation of self-directed learning in the Royal Australian College of General Practitioners Training Program*. South Australia: School of Medicine, Flinders University.
- Prochaska, J., & di Clemente, C. (1983). Stages and processes of self-change in smoking: Towards an integrative model of change. *Journal of Consulting and Clinical Psychology*, 51, 390–395.
- Regehr, G., Hodges, B., Tiberius, R., & Lofchy, J. (1996). Measuring self-assessment skills: An innovative relative ranking model. *Academic Medicine*, 71(10 suppl), S52–S54.
- Reiter, H. (2002). Self and peer assessment in tutorials: Application of a relative-ranking model. *Academic Medicine*, 77(11), 1134–1139.

- Rethans, J.-J., Norcini, J., Barón-Maldonado, M., Blackmore, D., Jolly, B. C., LaDuca, T., et al. (2002). The relationship between competence and performance: Implications for assessing practice performance. *Medical Education*, 36, 901–909.
- Rethans, J.-J., van Leeuwen, Y., Drop, R., van der Vleuten, C., & Sturmans, F. (1990). Competence and performance: Two different concepts in the assessment of quality of medical care. *Family Practice*, 7(3), 168–174.
- Richardson, I. M., Howie, J. G. R., & Berkeley, J. S. (1974). A further study of trainee general practitioners. *Journal of the Royal College of General Practitioners*, 24, 661–665.
- Risucci, D. A., Tortolani, A. J., & Ward, R. J. (1989). Ratings of surgical residents by self, supervisors and peers. *Surgery, Gynecology & Obstetrics*, 169, 519–526.
- Roter, D. L., Hall, J. A., & Aoki, Y. (2002). Physician gender effects in medical communication: A meta-analytic review. *Journal of the American Medical Association*, 288(6), 756–764.
- Royal Australian College of General Practitioners. (1974). *The RACGP Annual Report*. Melbourne: The Royal Australian College of General Practitioners.
- Royal Australian College of General Practitioners. (1998). *Five year strategic plan for evaluation of outcomes of the Royal Australian College of General Practitioners training program*. Melbourne: Royal Australian College of General Practitioners: Outcomes Evaluation Unit.
- Royal Australian College of General Practitioners. (1999a). *Making sense of GP learning: Companion to the Training Program Curriculum*. Melbourne: Royal Australian College of General Practitioners,.
- Royal Australian College of General Practitioners. (1999b). *Training Program Curriculum* (2nd ed.). Melbourne: The Royal Australian College of General Practitioners.
- Royal Australian College of General Practitioners. (2000). *Training Program handbook 2000* (1st ed.). Melbourne: The Royal Australian College of General Practitioners.
- Royal Australian College of General Practitioners. (2002). *Application to the Australian Medical Council for accreditation of its general practice education standards and processes*. Melbourne: Royal Australian College of General Practitioners.

- Russell, B. (1962). *History of western philosophy*. London: George Allen & Unwin.
- Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine*, 65(10), 611–621.
- Schmidt, M. R. (2000). You know more than you can say: In memory of Donald A Schön. *Public Administration Review*, 60(3), 266–274.
- Schön, D. (1987). *Educating the reflective practitioner*. San Francisco: Jossey-Bass.
- Sclabassi, S. E., & Woelfel, S. K. (1984). Development of self-assessment skills in medical students. *Medical Education*, 84, 226–231.
- Shapiro, J., & Talbot, Y. (1991). Applying the concept of the reflective practitioner to understanding and teaching family medicine. *Family Medicine*, 23, 450–456.
- Siahpush, M., Borland, R., & Scollo, M. (2002). Prevalence and socio-economic correlates among lone mothers in Australia. *Australian and New Zealand Journal of Public Health* 26, 132–135.
- Smith, R. C., Mettler, J. A., Stoffelmayr, B. E., Lyles, J. S., Marshall, A. A., Van Egeren, L. F., et al. (1995). Improving residents' confidence in using psychosocial skills. *Journal of General Internal Medicine*, 10, 315–320.
- Southgate, L. (1994). Freedom and discipline: Clinical practice and the assessment of clinical competence. *British Journal of General Practice*, 44(379), 87–92.
- Southgate, L., Campbell, M., Cox, J., Foulkes, J., Jolly, B., McCrorie, P., et al. (2001). The General Medical Council's performance procedures: The development and implementation of tests of competence with examples of general practice. *Medical Education*, 35 (Suppl. 1), 20–28.
- Speechley, M., Dickie, G. L., Weston, W. W., & Orr, V. (1993). Changes in residents' self-assessed competences during a two-year family practice program. *Academic Medicine*, 68(2), 163–165.
- Speechley, M., Weston, W. W., Dickie, G. L., & Orr, V. (1994). Self-assessed competence: Before and after residency. *Canadian Family Physician*, 40, 459–464.
- Spike, N. (1999). Evolution or revolution: the redevelopment of the RACGP fellowship examination. *Focus on Health Professional Education: a Multi-Disciplinary Journal*, 1(1), 76–84.

- Spike, N. A., & Veitch, P. C. (1990). Analysis of the RACGP fellowship examination results. *Australian Family Physician*, 19(5), 767–775.
- Steven, I. D. (1991). A patient satisfaction questionnaire as a teaching and comparative audit tool. *Quality Assurance in Health Care*, 3(1), 41–49.
- Stewart, J., O'Halloran, C., Barton, R., Singleton, S., Harrigan, P., & Spencer, J. (2000). Clarifying the concepts of confidence and competence to produce appropriate self-evaluation measurement scales. *Medical Education*, 34(11), 903–909.
- Stott, N., & Davis, R. (1979). The exceptional potential in each primary care consultation. *Journal of the Royal College of General Practitioners*, 29(201), 201–205.
- Stufflebeam, D., Foley, W. J., Gephart, W. J., Guba, E. J., Hammond, R. L., Merriman, H. O., et al. (1971). *Educational evaluation and decision making*. Illinois: Peacock Publishing.
- Tamblyn, R. (1994). Is the public being protected? Prevention of suboptimal medical practice through training programs and credentialing examinations. *Evaluation and the Health Professions*, 17(2), 198–221.
- Thomas, M., & Snadden, D. (2000). The 28-month scheme in general practice vocational training...going the distance? An interview study of trainers and GP registrars in Tayside. *Education for General Practice*, 11(1), 21–27.
- Thomson, A. (1990). An assessment of postgraduate examination of competence in general practice: Validity (Part 2). *New Zealand Medical Journal*, 103(889), 217–219.
- van de Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence. *Medical Education*, 39, 309–317.
- van Weyden, M. B. (2004). Medical education and hard science. *Medical Journal of Australia*, 181(6), 601.
- Victorian Auditor-General's Office (Ed.). (2003). *Improving literacy standards in government schools* (Vol. 26): Victorian Auditor-General's Office.
- Waitzkin, H. (1991). *The politics of medical encounters: How doctors and patients deal with social problems*. New Haven: Yale University Press.
- Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring self-assessment: Current state of the art. *Advances in Health Sciences Education*, 7(1), 63–80.
- Wearne, S. (2000). Pilot study on the factors that influence learning by general practitioners in central Australia. *Rural and Remote Health*, 3(223), 1–9.

- Wellard, R., & Edwards, H. (1999). Curriculum models for beginning practitioners. In J. Higgs and H. Edwards (Eds.), *Educating beginning practitioners: challenges for health professional education* (pp. 303). London: Butterworth Heinemann.
- Weller, D., & Dunbar, J. (2004). *History, policy and context*. Canberra: Australian Government Publishing Service.
- Westberg, J., & Jason, H. (1994). Fostering learners' reflection and self-assessment. *Family Medicine*, 26(5), 278–282.
- Wilde, S. (1998). *Twenty-five years under the microscope: A history of the RACGP Training Program 1973–1998*. South Melbourne: Royal Australian College of General Practitioners.
- Williams, P. (1998). Using theories of professional knowledge and reflective practice to influence educational change. *Medical Teacher*, 20(1), 28–34.
- Willis, E. (1989). *Medical dominance: the division of labour in Australian health care*. Sydney: Allen & Unwin.
- Willoughby, L. T., Arnold, L., & Calkins, V. (1981). Personal characteristics of medical students from urban and nonurban areas. *Journal of Medical Education*, 56(9), 717–726.
- Wooliscroft, J. O., Tenhaken, J., Smith, J., & J, C. (1993). Medical students clinical self-assessments: Comparisons with external measures of performance and the students' self-assessments of overall performance and effort. *Academic Medicine*, 68(4), 285–294.

Appendix A

Appendix A 1: Summary of self-assessment studies

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Antonelli, M. A. S. (1997). Accuracy of second-year medical students' self-assessment of clinical skills. <i>Academic Medicine</i> , 72(10), S63–65.	Validation study of knowledge and attitude data.	Second-year students in West Virginia University School of Medicine in Introduction to Clinical Medicine Course (n=87).	Physical diagnosis checklist, multiple choice questions on knowledge and global assessments of supervisors during clinical encounter.	High	Self-assessment of course performance (r=.49). Self-assessment and preceptor performance with preceptor grade (r=.31). Course grade and preceptor grades r=.24.
Arnold, L., Willoughby, T. L., and Calkins, V. E. (1985). Self-evaluation in undergraduate medical education: a longitudinal perspective. <i>Journal of Medical Education</i> , 60(1), 21–28.	Educational evaluation of longitudinal cohort study.	Baccalaureate medical students across four years in Missouri, Kansas City (n=211).	Global assessment on rating form with 11 attributes on a five-point scale.	High	Rating increased slightly. No associations between gender, ethnicity, and outcomes, but a moderate relationship found between locality and outcomes. Docent ratings were the best predictor of outcomes.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Biro, F. M., Siegel, D. M., Parker, R. M., and Gillman, M. W. (1993). A comparison of self-perceived clinical competencies in primary care residency graduates. <i>Pediatric Research</i> , 34(5), 555–559.	Educational evaluation using a cohort study	Primary care graduates compared to specialist graduates (n=178)	Global assessment of comfort with clinical vignette	High	χ^2 analysis showed that family medicine graduates are less confident in neonatal medicine and internal medicine. They are more confident in adolescent health care.
Bligh, J., and Slade, P. (1996). A questionnaire examining learning in general practice. <i>Medical Education</i> , 30(1), 65–70.	Educational evaluation with semi-structured interview about influences on learning	Primary care trainees (n=15).	Global judgment	High	Factor-based scales produced the following mean scores: learning from patients (69.8), openness to criticism (24.6), negative attitudes (19.7), need for guidelines (16.1), peer support (18.1) and academic approach (21.5).

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Blumenthal, D., Gokhale, M., Campbell, E., and Weissman, J. (2001). Preparedness for clinical practice: reports of graduating residents at academic health centers. <i>Journal of the American Medical Association</i> , 286(9), 1027–1033.	Educational evaluation using a cross-sectional survey	National survey of family medicine residents in internal medicine, paediatrics, family practice, obstetrics, gynaecology, surgery, psychiatry, and anaesthesiology (n=2526).	Checklist	High	Most residents viewed themselves as prepared by their courses, but as many as 10 percent in each speciality view themselves as under prepared.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Calhoun, J. G., Ten Haken, J., & Wooliscroft, J. O. (1990). Medical students' development and self- and peer-assessment skills: a longitudinal study. <i>Teaching and Learning in Medicine</i> , 2(1), 25–29.	Intervention to teach self-assessment	First and second year students at University of Michigan divided into two groups of three (n=236). Each student played the physician, patient and cameraperson. Faculty members scored the videotapes using checklists for a physical examination course.	Criterion-based	High	ANOVA used to examine the difference between students and peers. Fourth-year students were more critical of both their own and peers' performances compared to second-year students. Students' performance was dependent on the type of examination performed. Scheffé analysis showed that students' evaluation of their own and peers' performances mirrored faculties by the end of the clinical phase of training.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Clack, G. B. (1994). Medical graduates evaluate the effectiveness of their education. <i>Medical Education</i> , 28, 418-431.	Educational evaluation using cross-sectional survey.	Medical graduates two years after completion (n=371). To develop a curriculum, an expert group examined how prepared the graduates of Kings College school of Medicine and Dentistry are with regard to knowledge.	Self-report on level of factual content, skills imparted, and own attributes and skills.	High	Seventy percent of graduates reported having had a satisfactory education. Hospital graduates favoured more statistics and primary care physicians favoured behavioural science.
Clack, G. B., and Head, J. O. (1999). Gender differences in medical graduates' assessment of their personal attributes. <i>Medical Education</i> , 33(2), 101-105.	Educational evaluation using a cohort study	Five cohorts of doctors who qualified in King's College School of Medicine.	Participants were asked to indicate whether they had developed particular personal attributes by the time they had their medical qualifications.	High	Men reported being better equipped for leadership potential, and tolerance of ambiguity and uncertainty. Women reported being better equipped for being able to listen, work in a team, show a caring and compassionate nature and have satisfactory interpersonal relationships in professional life.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Cochran, S., and Spears, M. (1980). Student self-assessment and instructors' ratings: a comparison. <i>Journal of the American Dietetic Association</i> , 76(3), 253–257.	Intervention to teach self-assessment to compare performance against established criteria	Undergraduate dietetics students at the Kansas State University in a competency based program.	Instrument using critical incident technique to evaluate clinical performance. Analysis of grade point average.	High	Concurrence of the percentage of student and instructor rating improved as training progresses from the first observation to the third.
Donnelly, J. F., Boltri, J. M., Weeks, M. E., and Mengel, M. B. (1994). Measuring self-reported comfort with general family practice skills during a required third-year family practice clerkship. <i>Family Medicine</i> , 26(1), 21–26.	Educational evaluation using pre- and post test	Third-year medical students (n=179) in the Department of Family Practice and Community Medicine.	Global judgment on items such as interviewing, physical examination, clinical decision-making and management in the ambulatory setting, preventive medicine, assessment of psychosocial issues, and ethical problem solving.		Four factors explained 64.4 percent of the variance: relationships and values, history and physical, diagnosis and management and preventive medicine.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Falchikov, N., and Boud, D. (1989). Student self-assessment in higher education: a meta-analysis. <i>Review of Educational Research</i> , 59(4), 395–430.	Meta-analysis	Fifty-seven studies around a third are ‘wait and see’ studies, others measured students’ self-assessment skills, and only a few began from a theoretical standpoint.	Diverse	High	Correlations between teachers and students varied from –0.5 to 0.82. Quality of research design and the stage of training of students affected the results. Science produced better estimates than humanities.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Farnill, D., Hayes, S., and Todisco, J. (1997). Interviewing skills: self-evaluation by medical students. <i>Medical Education</i> , 31, 122–127.	Educational evaluation with validation of self-assessments through client and tutor assessments and use of videotaped performance.	Medical students carried out a self-assessment of their performance in two psychosocial interviews. They were videotaped and provided with feedback from clients, students and tutors. Students established a future learning plan.	Agreed criteria	High	Students' mean rating of 3.2 (adequate plus) was lower than that of the observers 3.6 (adequate to good). Correlations rarely exceeded $r=.30$ with overall self-assessment higher than usual ($r=.46$).

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Fincher, R.-M., Lewis, L. A., and Kuske, T. T. (1993). Relationships of interns' performances to their self-assessments of their preparedness for internship and to their academic performances in medical school. <i>Academic Medicine</i> , 68(2), S47–50.	Validation using comparison of cumulative grade average, self-assessments and program director ratings.	1990 and 1992 interns of the Medical College of the Georgia School of Medicine.	Rating items in areas of competency in basic science and clinical knowledge, history taking, physical examination, ordering tests, appropriate management, continuity of care, and willingness to engage in lifelong learning, and display interpersonal communication.	High	Mean rating of overall competencies by directors 3.7 to 4.3 out of five-point Likert scale. Self-reports by interns were lower ranging from 3.4 to 4.0. The correlation between students and directors on preparedness to practice was $r=.58$.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Fitzgerald J, Gruppen L, White C. 2000. The influence of task formats on the accuracy of medical students' self-assessments. <i>Academic Medicine</i> 75(7), 737-41.	Current framework examined whether self-assessment is a generalized skill or dependent on other characteristics of the task.	Fourth-year medical students at the University of Michigan engaged in self-assessment over ten stations of clinical examination consisting of performance, cognitive tests and paper and pen tests).	Standardized patient examination scored with a checklist, and paper and pen examination of students' medical knowledge. After the exam, students estimated their scores from one to 100 on each station.	High	Correlations suggested similarity on both tasks. Actual scores ranged from 65.5 to 87.6. Results suggested that students assess themselves fairly accurately.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Fitzgerald JT, White CB, Gruppen LD. 2003. A longitudinal study of self-assessment accuracy. <i>Medical Education</i> 37, 645–649.	Current framework examines the stability of self-assessment accuracy over time to gain a better understanding of self-assessment and to develop an intervention to assist poor self-assessment.	Students in the University of Michigan Medical School graduating class of 1999 (n=163), 2000 (n=169) and 2001 (n=168) assessed their performance on classroom examinations. Each self-assessment was compared to actual performance over three years. Multivariate analysis used to examine the stability of the measures.	Standardized patient examination scores with a checklist, and a paper and pen examination of students' medical knowledge.	High	Stability of self-assessment compared favourably with actual performance over time. Results suggested that self-assessment accuracy is a fairly stable characteristic that could be influenced by task familiarity.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Fraser, R. C., McKinley, R. K., and Mulholland, H. (1994a). Consultation competence in general practice: establishing the face validity of prioritized criteria in the Leicester assessment package. <i>British Journal of General Practice</i> , 44, 109–113.	Validation study for examining competence elements of the Leicester Assessment Package designed for formative and summative assessment.	Questionnaire sent to randomised stratified sample of members of the United Kingdom Association of Course Organizers.	Questionnaire had a six-point scale.	High	Response rate was 73 percent. Ninety-nine percent approved or strongly approved of the categories of consultation competence. Results suggested that the Leicester Assessment Package can be adopted for summative assessment purposes.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Ginsburg S, Regehr G, Hatala R, McNaughton N, Frohna A, et al. 2000. Context, conflict, and resolution: a new conceptual framework for evaluating professionalism. <i>Academic Medicine</i> 75(Suppl. 10) S6–S11.	Current framework examining peer- and self-assessment literature	Literature review	Review article	High	Ginsburg examined the literature on peer evaluation and self-assessment. He also commented on the high correlation with examination scores, building upon Arnold's initial work on peer evaluation. Finally, Ginsburg supported the appropriateness of evaluation of group trends as opposed to individual assessments.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Girgis A, Sanson-Fisher RW, Walsh R. 2001. Preventive medicine and other interactional skills of General Practitioners, surgeons and physicians: perceived competence and endorsement of postgraduate training. <i>Preventive Medicine</i> , 32, 78–81.	Educational evaluation comparing differences in self-assessments across a range of communication domains in three groups of recent graduates and perceptions of the need for training and assessment.	Mail survey (n=867) on four aspects of communication: prevention, education, therapy and general. Response rate of 45 percent was defended through the literature.	Survey of preventive items.		Low self-assessed competence ratings on preventive items: GPs 5–39 percent, surgeons 38–67 percent, and physicians 31–51 percent. Agreement with training on prevention: GPs 80–90 percent, surgeons 48–69 percent and physicians 72–82 percent.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Goedhuys, J., and Leroy, D. (1990). Belgian general practitioners on their vocational training: a postal inquiry. <i>Scandinavian Journal of Primary Health Care</i> , 8(4), 249–253.	Educational evaluation using cohort study.	General Practitioners from Interuniversity Centre for Vocational Training in Leuven, Belgium (n=177).	Survey about self-confidence on a six-point scale	High	Three factors identified: working atmosphere; learning results relevant to current project; and identification with the role of the GP. Mean scores and reliability high.
Gordon, M. J. (1991). A review of the validity and accuracy of self-assessment in health professions training. <i>Academic Medicine</i> , 66(12), 763–769.	Review article	Review of the literature of studies from 1970 to 1990.	Various instruments taken from the literature.	Only studies of a high standard were included.	Self-assessment performance capacity was related to generalized self-attributions and not influenced by external feedback. Validity of self-assessment was found to be low to moderate. Gordon concluded that there was much scope for improving the training of self-assessment.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Gordon, M. J. (1992). Self-assessment programs and their implications for health professions training. <i>Academic Medicine</i> , 67(10), 672–679.	Review article examining literature on validation of self-assessment and experts or objective testing.	Analysis of 11 studies.		Only studies that meet the criteria associated with accurate self-assessment were included.	Benefits of self-assessment included improvement in morale, motivation and communication between teachers and students. Cognitive benefits included improvements in knowledge, skills, performance and self-assessment of performance.
Grol, R., Tielens, V., and Mekkink, H. (1985). Attitude changes in the vocational training of general practitioners. <i>Medical Education</i> , 10, 479–86.	Educational evaluation using cohort study.	General Practitioners at University of Nijmegen, Belgium (n=84) are divided into groups of 12 and receive the same training.	Checklist of attitudinal dimensions of general practice in the areas of consulting skills and other forms of knowledge.	High	Clear change in attitudes in the areas of demarcation of roles of General Practitioners and specialists, and management of psychosocial problems. Trainees also reported more competence in medical technology, and ability to cope with uncertainty as they progress.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Harris, M. M., and Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. <i>Personnel Psychology</i> , 43–62.	Meta-analysis of 64 studies.	Secondary analysis.	Rating format.	High	Correlations were reported between different raters. Peer and supervisor ($r=.62$), self and supervisor ($r=.35$), and self and peer ($r=.36$).
Henbest, R. J., and Fehresen, S. G. (1985). Preliminary study at the Medical University of South Africa on student self-assessment as a means of evaluation. <i>Journal of Medical Education</i> , 60(1), 66–67.	Intervention to teach self-assessment.	Final year students doing four weeks at the Department of Family Medicine at the Medical University of Southern Africa (n=19) participated in a self-assessment exercise.	Agreed criteria.	High	Small difference found between faculty mean score and student ratings. High correlation between faculty and student marks (.74).

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Hojat, M., Gonnella, J. S., Veloski, J. J., & Erdmann, J. B. (1996). Jefferson medical college longitudinal study: a prototype for evaluation of changes. <i>Education for Health</i> , 9(1), 99–113.	Review of longitudinal educational evaluation	Longitudinal analysis of database with 6700 students and graduates from Jefferson Medical College produced 53 published studies	Admissions data, US Medical Licensing Examination, Graduation questionnaire follow-up survey on specialties after training, competence ratings, and performance and licensing data.	High	Difference identified between students from different undergraduate institutions raising questions about the tests of these institutions. A background in basic science in medical school was found to contribute to later clinical performance. Gender differences were found with basic science examinations, anticipated income and choice of speciality. Some research on personal qualities in relation to achievement showed that personality measures were significant predictors of academic achievement. Graduates assigned high ratings to Jefferson for their career preparation.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Jansen, J. J. M., Tan, L. H. C., van der Vleuten, C. P. M., van Luijk, S. J., Rethans, J. J., and Grol, R. P. T. M. (1995). Assessment of competence in technical clinical skills of general practitioners. <i>Medical Education</i> , 29, 247–253.	Validation study of performance tests, self-assessments, and knowledge.	Comparison of GPs (n=49) and general practice trainees (n=47).	Checklist on procedures performed, written knowledge test, and self-assessment questionnaire on a five-point scale.	High	Low correlations reported between self-assessment scales and performance-based tests.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Jansen, J. J. M., Grol, R. P. T. M., Crebolder, H. F. J. M., Rethans, J.-J., and van der Vleuten, C. P. M. (1998). Failure of feedback to enhance self-assessment skills of general practitioners. <i>Teaching and Learning in Medicine</i> , 10(3), 145–151.	Intervention to teach self-assessment. Validation comparing self-assessment score and knowledge test, and performance test between control and intervention group.	GPs who received skills training at three and six months (n=60).	Knowledge test, performance based self-assessment questionnaire.	High	At three months, mean scores on self-assessment test and knowledge test increased.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Keck, J. W., and Arnold, L. (1979). Efficacy of cognitive/noncognitive measures in predicting resident-physician performance. <i>Journal of Medical Education</i> , 54(10), 759–765.	Validation study comparing evaluation form data with National Board of Medical examiners Part III examination.	Study of supervisors and final year Bachelor of Arts and medical degree program (n=66).	Evaluation form with seven-point scales measuring behaviour such as professional responsibility, relations with colleagues, self-appraisal skills, relations with patients, critical thinking ability and problem solving.	High	Grade point average correlated with clinical performance ($r=.34$). Internal medicine instructors' evaluations also correlated with performance ($r=.26$), and year six evaluations ($r=.31$). Non-cognitive and cognitive variables functioned better as predictors of performance ($R^2=.37$).

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Kegel-Flom, P. (1975). Predicting supervisor, peer, and self-ratings of intern performance. <i>Journal of Medical Education</i> , 50, 812–815.	Validation study of knowledge, performance and personality inventories.	Entire graduate class of 1968, University of California, School of Medicine (n=110).	Medical College and Admission test scores, interview ratings, personality inventories, Medical Grade School Average and self-ratings on dimensions of knowledge, clinical effectiveness, and interpersonal differences.	High	Supervisors rated interns higher than self- and peer-ratings. Regression results showed that supervisors rate interns higher who do well in science and excelled in medical course work. The best combination to predict self-ratings was the Medical Grade School Average, psychological inventory (positive) and the flexibility scale (negative). Those who rated themselves highly were forceful but rigid and achieved high academic results. Personality was the best predictor of peer ratings.
Kolm, P., and Verhulst, S. J. (1987). Comparing self- and supervisor evaluations: a different view. <i>Evaluation and Health Professions</i> , 10(1), 80–89.	Validation study of students and supervisor ratings.	Interns from the Southern Illinois School of Medicine. Data from 151 of 397 residents in the graduating classes of 1977 and 1982.	Evaluation form included history taking, problem identification, diagnostic approach, management, learning habits, and relationships with patients.	High	Correlations ranged from $r = -.02$ to $.26$.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Linn, B. S., Arostegui, M., and Zeppa, R. (1975). Performance rating scale for peer and self-assessment. <i>British Journal of Medical Education</i> , 9, 98–101.	Validation study of students and ten peers.	University of Miami School of Medicine (n=54).	Some of the items were fund of knowledge, conscientiousness, leadership, clinical judgment, doctor-patient relationship, team spirit, interpersonal relationships, intellectual curiosity, and technical ability.	High	Reasonable reliability on test and re-test. Students rated themselves lower. Peer ratings toward above average. Students were interested in the process of self-assessment.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Martin, D., Regehr, G., Hodges, B., and McNaughton, N. (1998). Using videotaped benchmarks to improve the self-assessment ability of family practice residents. <i>Academic Medicine</i> , 73(11), 1201–1206.	Intervention to teach self-assessment.	Family practice residents (n=50) performed a patient interview at the University of Toronto. After the interview, the resident and two experts evaluated the resident.	Agreed criteria with students using vignettes to evaluate own performance.	High	Correlations between resident and experts immediately after interview were moderate ($r=.38$) and increased ($r=.52$) after residents watched the videotaped vignettes.
Morton JB, Macbeth WAAG. 1977. Correlations between staff-, peer- and self-assessments of fourth-year students in surgery. <i>Medical Education</i> , 11(3), 167–70.	Validation study of staff and student assessments.	Surgery students in New Zealand (n=138). Staff assessment comprises a short answer paper, supervisor marks and ratings of students' presentation, and continuous assessment.	Mean score of four staff.	High	Students found it difficult to assess themselves. Mean scores for student assessments were lower than peer assessments. Students with failing self-assessments were found to have psychological problems. High correlations were found between peer and staff assessments.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Regehr, G., Hodges, B., Tiberius, R., and Lofchy, J. (1996). Measuring self-assessment skills: an innovative relative ranking model. <i>Academic Medicine</i> , 71(10 Suppl.), S52–S54.	Current framework using relative ranking to evaluate students' strengths and weaknesses and compare to examiner ratings.	Third year clinical clerks on a six-week psychology rotation in Toronto (n=25) complete an instrument based on review using a standardized patient. Examiner and student completed the rating form.	Rating form with a scale ranging from 1 to 10 (1=needs more work and 10=needs least work).	High	Mean score of student assessment rankings was moderate when corrected for expert unreliability (mean 0.58). There was a wide variation of abilities of students in assessing themselves. Correlation between student self-assessment and experts overall rating was $r=0.19$.
Reiter, H. (2002). Self and peer assessment in tutorials: application of a relative-ranking model. <i>Academic Medicine</i> , 77(11), 1134–1139.	Current framework.	Students in McMaster University of Health Sciences (problem based curriculum) completed ranking forms with their tutors and two of the students' peers.	Checklist	High	Poor correlations found for self-and peer-ratings and self- and tutor-ratings). Inter rater reliability also poor for tutor-peer assessments. Results did not improve over time.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Risucci, D. A., Tortolani, A. J., and Ward, R. J. (1989). Ratings of surgical residents by self, supervisors and peers. <i>Surgery, Gynecology & Obstetrics</i> , 169, 519–526.	Validation study using factor analysis.	Comparison between various raters (surgeons n=61), peers n=32) and American Board of Surgery In Training	Rating form	High	Overall ratings by peers and supervisors highly correlated ($r=.92$). Average of overall ratings by peers and supervisors correlate moderately with raw score ($r=.58$).

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Sclabassi, S. E., and Woelfel, S. K. (1984). Development of self-assessment skills in medical students. <i>Medical Education</i> , 84, 226–231.	Validation study using self-assessment activity	Third-year medical students (n=36) completed an evaluation checklist with their supervisors.	Evaluation checklist on a five-point scale. The scale ranged from outstanding to inadequate in the areas of history taking and physical examinations, motivation and ability to accept criticism, interpersonal relationships, problem solving ability, overall clinical performance and integrity.	High	Five percent of the students' forms were in agreement with instructors when all of the items on the students self-assessment form were tallied.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Smith, R. C., Mettler, J. A., Stoffelmayr, B. E., Lyles, J. S., Marshall, A. A., Van Egeren, L. F., et al. (1995). Improving residents' confidence in using psychosocial skills. <i>Journal of General Internal Medicine</i> , 10, 315–320.	Randomised control trial to test the effectiveness of educational intervention in psychosocial skills	Twenty-six first year students in internal medicine and family practice residents at the University of Michigan are assigned to a control group or a one-month course in psychosocial skills in primary care (n=26).	Two pre- and post surveys completed twice.	High	Trained residents reported higher levels of confidence in using psychosocial skills. They anticipated better outcomes in the areas of facilitative patient communication, and were more committed to being emotionally sensitive. They also demonstrated more knowledge of psychosocial medicine than those who did not participate in the course.
Speechley, M., Dickie, G. L., Weston, W. W., and Orr, V. (1993). Changes in residents' self-assessed competences during a two-year family practice program. <i>Academic Medicine</i> , 68(2), 163–165.	Educational evaluation	Residents entering the University of Western Ontario Faculty of Medicine were invited to participate and 42 agreed. One group started in a hospital and one group in family medicine.	The residents completed at baseline and after 6, 12 and 24 months on a scale based on the International Classification for Health Problems (ICHPPC-2).	High	At baseline, mean scores between hospital and family practice residents were the same. At six months, family practice residents rated themselves as more competent with management of skin problems, other clinical situations and technical skills. Hospital residents report more competence with management of accidents. No differences at two years.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Speechley, M., Weston, W. W., Dickie, G. L., and Orr, V. (1994). Self-assessed competence: before and after residency. <i>Canadian Family Physician</i> , 40, 459–464.	Education evaluation using pre- and post test	Residents entering the University of Western Ontario Faculty of Medicine were invited to participate (n=42). One group started in a hospital and one group in family medicine.	The residents completed at baseline and after 6, 12 and 24 months a questionnaire based on the International Classification for Health Problems (ICHPPC-2).	High	Twenty percent of residents reported a lack of confidence in 47 of 177 area of competence. Self-assessed competence increased as residents progressed through the program.
Stewart, J., O'Halloran, C., Barton, R., Singleton, S., Harrigan, P., and Spencer, J. (2000). Clarifying the concepts of confidence and competence to produce appropriate self-evaluation measurement scales. <i>Medical Education</i> , 34, 903–909.	Current framework with literature review	Not applicable	Review examining the meaning and significance of competence and confidence.	High	Concluded that confidence and self-assessed competence were not the same. Positive experience of confidence related to house officers' experience of competence, but negative experience of confidence might relate to anxiety. Some congruence between interview and response to the instrument was found.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Thomas, M., and Snadden, D. (2000). The 18-month scheme in general practice vocational training...going the distance? An interview study of trainers and GP registrars in Tayside. <i>Education for General Practice</i> , 11, 21–27.	Educational evaluation using qualitative analysis.	Qualitative study using interviews with trainers (n=5) and GPs (n=6) interviewed.	Interviews about overall value of training, exposure to general practice and learning outside the practice.	High	Trainees were generally positive about their induction into general practice, development of autonomy, the absence of exams and focus on real-world experience. There were some concerns about learning outside the practice, and missing the hospital experience, isolation, access to day release and the timeliness of exam content. Trainers' issues were around practice-based learning, the training relationship, learning outside the practice, and having a creative vision for general practice.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Tracey, J. M., Arroll, B., Richmond, D. E., and Barham, P. M. (1997). The validity of general practitioners' self-assessment of knowledge: cross sectional study. <i>British Medical Journal</i> , 315, 1426–1428.	Validation study with cross-sectional survey comparing trainees with non-trainees.	Cross sectional mailed study with New Zealand General Practitioners (n=67). Some are in the training program others are not.	Knowledge test	High	Correlations between test and self-assessment were low in areas such as thyroid disorders ($r=0.19$), non-insulin diabetes ($r=0.21$) and sexually transmitted diseased ($r=0.19$).
Ward M, Gruppen L, Regehr G. 2002. Measuring self-assessment: current state of the art. <i>Advances in Health Sciences Education</i> , 7(1), 63-80.	Current framework in this literature review compared the methodologies of other studies on self, and supervisors.	Literature review	Review article	High	Studies showed poor results in the accuracy of self-assessment with methodological issues affecting their design. Ward et al. recommend that self-assessment should be conceptualised so as to help physicians identify strengths and weaknesses.

Published study	Method	Subjects and level of assessment	Instrument	Strength of design	Results
Wooliscroft JO, Tenhaken J, Smith J, J C. 1993. Medical students' clinical self-assessments: comparisons with external measures of performance and the students' self-assessments of overall performance and effort. <i>Academic Medicine</i> 68(4), 285–294.	Validation study.	Third-year students at the University of Michigan hospital in 1998 (n=142).	Questionnaire on abilities in clinical skills, discharge of patient care, and clinical knowledge. External data college-point grade averages, exams, and faculty ratings.	High	Relatively weak to moderate correlations between external and student self-assessments.

Appendix B

Appendix B 1: Gender by Domain one	235
Appendix B 2: Locality by Domain one	235
Appendix B 3: Gender by Domain two	235
Appendix B 4: Gender by Domain three	236
Appendix B 5: Locality by Domain four	236
Appendix B 6: Gender by Domain five	237
Appendix B 7: Gender by particular patient groups	238
Appendix B 8: Locality by particular patient groups	239
Appendix B 9: Stage of training by particular patient presentations	240
Appendix B 10: Stage of training by cardiovascular to dermatological problems	241
Appendix B 11: Stage of training by musculoskeletal to reproductive problems	242
Appendix B 12: Stage of training by neurological to pediatrics	243
Appendix B 13: Stage of training by HIV/AIDS to preventive medicine	244
Appendix B 14: Stage of training by public health medicine	245
Appendix B 15: Gender by particular patient presentations	246
Appendix B 16: Gender by major trauma and disasters to psychological problems	247
Appendix B 17: Gender by neurological problems to preventive medicine	248

Appendix B 34: Workload by satisfaction.....	265
Appendix B 35: Workload by further satisfaction items	266
Appendix B 36: Longitudinal findings for satisfaction items.....	267
Appendix B 37: Longitudinal findings for satisfaction with information items	268

Appendix B 1: Gender by Domain one

	Scale scores Mean* (Standard deviation)		Statistical test		
	Male	Female	<i>t</i> test	<i>df</i>	Sig.
Communicate with patients effectively					
2000	98.2 (8.8)	99.5 (4.1)	-2.0	288	<i>p</i> <.05

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 2: Locality by Domain one

	Scale scores Mean* (Standard deviation)		Statistical test		
	Urban areas	Rural and remote areas	<i>t</i> test	<i>df</i>	Sig.
Provide continuity of care using a whole-person approach					
2000	93.9 (14.5)	96.5 (11.0)	-2.0	361	<i>p</i> <.05

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 3: Gender by Domain two

	Scale scores Mean* (Standard deviation)		Statistical test		
	Male	Female	<i>t</i> test	<i>df</i>	Sig.
Order cost-effective investigations					
1999	88.8 (19.3)	88.5 (20.3)	2.0	526	<i>p</i> <.05
2001	92.8 (15.7)	87.6 (18.7)	3.3	458	<i>p</i> <.01

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 4: Gender by Domain three

	Scale scores Mean* (Standard deviation)		Statistical test		
	Male	Female	<i>t</i> test	<i>df</i>	Sig.
Undertake a public health role					
2001	84.0 (19.8)	88.5 (19.2)	-1.4	375	<i>p</i> <.05

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 5: Locality by Domain four

	Scale scores Mean* (Standard deviation)		Statistical test		
	Urban areas	Rural and remote areas	<i>t</i> test	<i>df</i>	Sig.
Engage in ongoing, self-directed learning and professional development					
2000	92.9 (16.1)	96.3 (11.9)	-2.4	365	<i>p</i> <.05

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 6: Gender by Domain five

	Scale scores		Statistical test		
	Mean* (Standard deviation)				
	Male	Female	<i>t</i> test	<i>df</i>	Sig.
Effectively maintain and use patient records					
1999	90.3 (18.9)	93.4 (15.8)	-2.2	440	$p<.05$
Implement effective practice management systems					
1999	67.0 (26.0)	62.8 (24.7)	2.9	490	$p<.01$
2000	71.8 (25.8)	64.8 (24.0)	3.2	446	$p<.01$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 7: Gender by particular patient groups

	Scale scores Mean* (Standard deviation)		Statistical test		
	Male	Female	<i>t</i> test	<i>df</i>	Sig.
Infants					
1999	83.3 (25.7)	92.9 (16.1)	-5.4	443	<i>p</i> <.001
Children					
1999	91.8 (18.8)	96.8 (10.4)	-4.0	410	<i>p</i> <.001
Women					
1999	94.7 (14.1)	97.2 (10.9)	-2.5	510	<i>p</i> <.01
2000	95.0 (13.5)	98.5 (7.3)	-3.5	310	<i>p</i> <.01
2001	94.5 (14.9)	99.4 (4.5)	-4.5	227	<i>p</i> <.001
Men					
2000	99.1 (6.3)	92.4 (15.4)	7.0	453	<i>p</i> <.01
2001	98.7 (6.5)	94.7 (13.5)	4.2	402	<i>p</i> <.001
Neonates					
1999	67.4 (28.9)	74.5 (27.1)	-3.1	611	<i>p</i> <.01
Older people					
2000	98.0 (8.5)	94.8 (13.5)	3.4	535	<i>p</i> <.01
2001	97.5 (10.0)	95.2 (12.4)	2.2	461	<i>p</i> <.05

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 8: Locality by particular patient groups

	Scale scores Mean* (Standard deviation)		Statistical test		
	Urban areas	Rural and remote areas	<i>t</i> test	<i>df</i>	Sig.
People from rural remote communities					
2000	89.2 (19.4)	82.4 (21.6)	3.8	499	<i>p</i> <.01
2001	90.3 (18.7)	84.7 (21.1)	3.2	450	<i>p</i> <.01
Adolescents					
2000	93.1 (14.5)	96.5 (11.7)	-2.5	345	<i>p</i> <.05
Women					
2000	95.5 (12.9)	97.9 (9.1)	-2.1	373	<i>p</i> <.05
Men					
2000	93.4 (14.9)	96.2 (10.6)	-2.2	371	<i>p</i> <.05
Indigenous people					
2000	75.8 (24.3)	84.3 (20.9)	-3.6	330	<i>p</i> <.01
2001	80.1 (22.0)	84.7 (22.5)	-2.1	443	<i>p</i> <.05
People from rural remote communities					
2000	81.5 (22.2)	92.5 (15.1)	-5.8	377	<i>p</i> <.01
2001	83.1 (22.3)	94.6 (12.9)	-6.9	433	<i>p</i> <.001

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 9: Stage of training by particular patient presentations

	Scale scores					Statistical test		
	Mean* (Standard deviation)							
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
Eye problems								
1999	62.3 (27.5)	69.9 (25.0)	84.4 (20.0)	85.5 (20.9)	90.0 (18.8)	34.8	4	<i>p</i> <.001
2000	65.5 (25.0)	80.0 (24.3)	83.1 (21.3)	90.9 (16.0)	87.4 (21.4)	21.3	4	<i>p</i> <.001
2001	63.0 (24.8)	77.5 (22.3)	80.9 (22.1)	86.3 (20.2)	90.6 (17.2)	21.8	4	<i>p</i> <.001
Ear, nose and throat problems								
1999	69.3 (26.7)	84.5 (19.9)	94.9 (13.9)	96.0 (11.6)	95.9 (12.7)	52.4	4	<i>p</i> <.001
2000	74.9 (25.8)	88.1 (20.2)	93.3 (14.7)	95.1 (11.8)	96.1 (12.0)	26.3	4	<i>p</i> <.001
2001	69.1 (26.6)	91.9 (16.5)	87.0 (20.9)	96.4 (11.3)	96.4 (11.6)	36.7	4	<i>p</i> <.001
Respiratory problems								
1999	94.4 (13.2)	96.5 (10.2)	99.6 (3.5)	99.2 (5.1)	99.2 (6.2)	8.5	4	<i>p</i> <.001
2000	95.5 (12.5)	98.7 (6.4)	98.5 (9.2)	99.8 (2.9)	99.3 (4.7)	5.4	4	<i>p</i> <.001
2001	93.8 (17.2)	98.2 (7.6)	99.4 (4.5)	99.5 (4.1)	99.8 (2.4)	8.5	4	<i>p</i> <.001

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ¶Subsequent trainees ||Former trainees

Appendix B 10: Stage of training by cardiovascular to dermatological problems

	Scale scores Mean* (Standard deviation)					Statistical test		
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
Cardiovascular problems								
1999	90.4 (18.7)	90.4 (16.4)	96.7 (10.0)	95.5 (13.6)	96.8 (11.1)	6.6	4	<i>p</i> <.001
2000	90.3 (16.8)	93.3 (16.3)	95.4 (13.0)	97.3 (9.1)	97.5 (9.6)	6.0	4	<i>p</i> <.001
2001	88.9 (20.5)	91.9 (18.3)	94.4 (14.1)	96.1 (10.7)	99.0 (5.3)	9.5	4	<i>p</i> <.001
Gastrointestinal and digestive problems								
1999	91.5 (16.5)	93.0 (15.0)	96.7 (10.0)	96.3 (11.3)	97.8 (9.0)	5.9	4	<i>p</i> <.001
2000	91.0 (17.2)	93.7 (13.2)	95.9 (12.5)	98.5 (8.0)	98.2 (8.5)	8.0	4	<i>p</i> <.001
2001	88.1 (20.8)	98.2 (7.6)	98.1 (10.1)	98.5 (7.0)	99.0 (6.7)	13.5	4	<i>p</i> <.001
Dermatological problems								
1999	54.7 (24.3)	62.9 (24.9)	79.4 (22.8)	84.9 (20.1)	88.2 (20.0)	54.5	4	<i>p</i> <.001
2000	55.1 (25.2)	67.9 (22.6)	80.5 (21.2)	83.6 (20.7)	89.6 (20.1)	43.1	4	<i>p</i> <.001
2001	54.9 (24.4)	66.7 (26.0)	72.8 (24.3)	86.3 (19.4)	89.8 (17.9)	38.9	4	<i>p</i> <.001

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ¶Subsequent trainees ||Former trainees

Appendix B 11: Stage of training by musculoskeletal to reproductive problems

	Scale scores					Statistical test		
	Mean* (Standard deviation)					ANOVA F	df	Sig.
	H†	B‡	A§	S	F¶			
Musculoskeletal problems								
1999	75.6 (28.9)	74.8 (24.8)	83.0 (20.8)	84.8 (21.0)	86.9 (20.0)	7.7	4	p<.001
2000	70.5 (25.0)	76.6 (23.2)	84.6 (22.1)	89.0 (18.6)	89.6 (19.8)	16.2	4	p<.001
2001	71.6 (26.2)	74.8 (27.7)	85.8 (19.0)	89.9 (18.4)	88.5 (19.8)	10.9	4	p<.001
Genitourinary problems								
1999	83.6 (21.8)	88.4 (19.3)	95.6 (12.4)	95.6 (10.2)	96.1 (11.4)	16.8	4	p<.001
2000	85.0 (20.7)	89.5 (17.5)	93.8 (15.5)	96.8 (11.4)	96.8 (9.8)	12.8	4	p<.001
2001	82.1 (24.0)	90.1 (17.3)	92.6 (14.0)	97.4 (9.9)	96.5 (11.3)	14.4	4	p<.01
Reproductive problems								
1999	74.5 (27.0)	80.3 (24.2)	89.1 (17.9)	91.3 (17.0)	94.9 (14.1)	22.6	4	p<.01
2000	76.9 (26.4)	80.3 (25.2)	90.8 (19.1)	93.9 (14.2)	93.7 (15.2)	17.2	4	p<.001
2001	73.5 (27.0)	89.2 (21.0)	82.1 (22.1)	92.0 (16.1)	93.4 (15.7)	14.7	4	p<.01

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ¶Subsequent trainees ||Former trainee

Appendix B 12: Stage of training by neurological to pediatrics

	Scale scores Mean* (Standard deviation)					Statistical test		
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
Neurological problems								
1999	71.6 (25.1)	74.8 (23.6)	80.0 (21.1)	84.1 (19.7)	81.3 (19.3)	8.5	4	<i>p</i> <.01
2000	74.2 (25.5)	77.5 (21.2)	82.6 (18.7)	85.8 (18.4)	85.6 (20.3)	6.6	4	<i>p</i> <.001
2001	70.4 (25.6)	77.5 (22.3)	76.5 (21.1)	84.8 (19.5)	88.0 (18.4)	10.5	4	<i>p</i> <.001
Sexually transmitted diseases								
1999	73.6 (26.3)	84.4 (21.3)	92.0 (16.7)	95.2 (11.7)	95.1 (13.4)	31.3	4	<i>p</i> <.01
2000	73.4 (25.7)	85.4 (21.1)	91.8 (17.7)	96.1 (11.5)	94.8 (13.9)	27.7	4	<i>p</i> <.001
2001	67.9 (28.9)	84.7 (23.0)	88.3 (20.6)	94.8 (14.1)	96.2 (11.2)	32.1	4	<i>p</i> <.001
Drug and alcohol problems								
1999	69.3 (27.0)	74.0 (24.2)	78.4 (23.0)	80.8 (22.9)	79.1 (23.5)	5.3	4	<i>p</i> <.01
2000	68.6 (25.5)	72.4 (25.1)	80.0 (22.7)	79.7 (21.5)	81.3 (21.7)	6.0	4	<i>p</i> <.001
2001	63.6 (27.7)	71.2 (22.4)	71.0 (25.1)	82.9 (21.3)	80.4 (22.2)	9.3	4	<i>p</i> <.001
Pediatrics								
1999	83.6 (21.8)	88.4 (19.2)	95.6 (12.4)	95.6 (10.2)	96.1 (11.4)	28.5	4	<i>p</i> <.001
2000	83.9 (23.6)	95.6 (12.3)	94.4 (13.9)	97.8 (8.3)	97.3 (10.7)	16.6	4	<i>p</i> <.01
2001	82.7 (24.9)	97.3 (9.2)	96.3 (10.6)	97.9 (9.1)	98.3 (8.9)	18.9	4	<i>p</i> <.001

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ¶Subsequent trainees ||Former trainees

Appendix B 13: Stage of training by HIV/AIDS to preventive medicine

	Scale scores					Statistical test		
	Mean* (Standard deviation)					ANOVA F	df	Sig.
HIV/AIDS problems								
1999	53.5 (24.1)	57.7 (25.5)	63.0 (24.1)	63.8 (25.3)	63.5 (25.2)	4.0	4	$p<.01$
2000	55.4 (24.1)	56.8 (23.5)	65.1 (25.3)	64.2 (24.0)	65.3 (24.9)	4.0	4	$p<.01$
Emergency medicine								
1999	85.5 (23.8)	84.1 (21.8)	90.8 (16.5)	85.7 (20.8)	82.2 (23.3)	2.4	4	$p<.05$
Psychological problems								
1999	95.1 (25.3)	79.7 (21.9)	81.7 (22.4)	89.7 (17.6)	86.5 (20.7)	8.8	4	$p<.001$
2000	74.9 (26.2)	79.7 (22.9)	91.3 (17.9)	87.9 (18.5)	91.0 (17.2)	12.5	4	$p<.001$
2001	72.2 (26.0)	84.7 (21.7)	80.2 (22.0)	91.7 (16.7)	89.8 (18.5)	12.0	4	$p<.001$
Preventive medicine								
1999	82.3 (22.7)	92.5 (14.7)	96.7 (10.1)	98.2 (7.7)	98.4 (8.0)	30.7	4	$p<.01$
2000	86.1 (20.6)	94.3 (14.2)	97.9 (8.1)	98.5 (8.0)	97.1 (10.9)	15.0	4	$p<.001$
2001	86.4 (22.0)	94.6 (12.5)	94.4 (12.5)	98.1 (7.6)	97.7 (8.4)	12.2	4	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ¶Subsequent trainees ||Former trainees

Appendix B 14: Stage of training by public health medicine

	Scale scores					Statistical test		
	Mean* (Standard deviation)							
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
Public health medicine								
1999	75.9 (24.0)	84.6 (20.8)	85.1 (20.0)	87.7 (19.6)	87.6 (19.1)	6.7	4	<i>p</i> <.001
2000	74.5 (25.1)	86.3 (20.0)	94.4 (15.1)	91.7 (15.6)	85.6 (19.1)	13.9	4	<i>p</i> <.001
2001	75.3 (22.6)	91.1 (15.0)	87.0 (18.8)	91.5 (16.3)	92.5 (15.1)	11.9	4	<i>p</i> <.001

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ¶Subsequent trainees ||Former trainees

Appendix B 15: Gender by particular patient presentations

	Scale scores Mean* (Standard deviation)		Statistical test		
	Male	Female	<i>t</i> test	<i>df</i>	Sig.
Eye problems					
1999	86.6 (21.5)	80.3 (23.6)	3.2	500	$p<.01$
2000	88.5 (19.4)	80.8 (22.9)	3.9	457	$p<.001$
Cardiovascular problems					
1999	95.5 (13.3)	93.1 (15.2)	2.1	555	$p<.05$
2000	96.8 (10.3)	94.1 (14.6)	2.6	539	$p<.05$
2001	97.3 (9.7)	95.0 (13.9)	2.2	460	$p<.05$
Musculoskeletal problems					
1999	84.3 (22.6)	76.3 (25.5)	2.5	606	$p<.01$
2000	89.4 (18.5)	78.8 (24.1)	5.8	532	$p<.01$
2001	90.3 (17.9)	82.1 (23.6)	4.3	463	$p<.001$
Emergency medicine					
1999	89.0 (19.9)	82.4 (22.8)	3.8	558	$p<.001$
2001	91.5 (18.0)	84.4 (21.7)	3.8	457	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 16: Gender by major trauma and disasters to psychological problems

	Scale scores		Statistical test		
	Mean* (Standard deviation)				
	Male	Female	<i>t</i> test	<i>df</i>	Sig.
Major trauma and disasters					
1999	77.1 (26.9)	66.1 (27.0)	4.9	509	<i>p</i> <.001
2000	76.9 (26.5)	62.0 (26.2)	6.5	539	<i>p</i> <.01
2001	79.2 (25.8)	66.5 (25.3)	5.2	424	<i>p</i> <.001
Reproductive problems					
1999	79.4 (23.7)	91.5 (18.5)	-6.7	426	<i>p</i> <.001
2000	81.9 (23.9)	92.4 (17.6)	-5.6	379	<i>p</i> <.01
2001	83.5 (21.7)	93.3 (17.0)	-5.3	367	<i>p</i> <.001
Sexually transmitted diseases					
1999	84.4 (22.0)	91.3 (18.2)	-4.0	442	<i>p</i> <.001
2000	84.6 (21.9)	92.8 (16.9)	-4.7	393	<i>p</i> <.01
2001	86.5 (22.2)	93.8 (16.1)	-4.0	348	<i>p</i> <.001
Psychological problems					
1999	79.0 (24.2)	85.6 (20.2)	-3.50	448	<i>p</i> <.001

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 17: Gender by neurological problems to preventive medicine

	Scale scores		Statistical test		
	Mean*				
	(Standard deviation)				
	Male	Female	<i>t</i> test	<i>df</i>	Sig.
Neurological problems					
2000	87.2 (18.8)	78.3 (22.0)	5.0	514	<i>p</i> <.01
2001	86.3 (20.4)	80.4 (21.3)	3.0	463	<i>p</i> <.01
Drug and alcohol problems					
2000	80.1 (23.0)	74.6 (23.8)	2.7	538	<i>p</i> <.01
Preventive medicine					
1999	83.7 (21.1)	95.1 (20.9)	-2.4	468	<i>p</i> <.05

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 18: Workload by particular patient presentations

	Scale scores					Statistical test		
	Mean* (Standard deviation)					ANOVA F	df	Sig.
	1†	2‡	3§	4¶	5			
Eye problems								
1999	74.0 (29.3)	77.2 (25.0)	79.9 (23.1)	88.3 (20.6)	84.8 (24.7)	4.7	4	<i>p</i> <.01
2000	77.8 (24.6)	77.9 (24.7)	86.9 (20.1)	91.7 (16.4)	76.0 (26.4)	8.9	4	<i>p</i> <.01
2001	78.0 (24.8)	81.8 (22.4)	87.6 (19.7)	87.7 (20.7)	89.1 (15.8)	3.5	4	<i>p</i> <.01
Ear, nose and throat problems								
1999	82.4 (22.7)	87.3 (20.6)	91.6 (18.0)	94.0 (15.4)	91.4 (18.7)	5.0	4	<i>p</i> <.01
2000	86.7 (22.9)	89.6 (18.9)	93.8 (15.1)	95.2 (13.0)	82.5 (23.2)	8.0	4	<i>p</i> <.01
2001	85.1 (21.9)	93.0 (16.3)	93.8 (15.0)	94.6 (15.9)	94.9 (14.0)	3.6	4	<i>p</i> <.01
Respiratory problems								
1999	95.6 (12.8)	97.8 (8.3)	99.1 (5.5)	98.2 (9.0)	99.1 (5.6)	2.7	4	<i>p</i> <.05
Cardiovascular problems								
1999	90.7 (18.1)	92.1 (17.0)	95.4 (12.0)	97.2 (10.6)	100.0 (0.0)	4.9	4	<i>p</i> <.01
Gastrointestinal problems								
1999	94.6 (13.6)	92.5 (15.9)	96.9 (9.6)	97.2 (10.6)	97.1 (9.5)	3.7	4	<i>p</i> <.01

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients ||161≤ patients

Appendix B 19: Workload by further patient presentations

	Scale scores Mean* (Standard deviation)					Statistical test		
	1†	2‡	3§	4¶	5	ANOVA F	df	Sig.
Pediatrics	89.7 (21.7)	93.4 (16.3)	96.6 (11.0)	97.2 (11.6)	96.2 (10.8)	4.2	4	p <.01
1999	89.7 (21.7)	93.4 (16.3)	96.6 (11.0)	97.2 (11.6)	96.2 (10.8)	4.2	4	p <.01
2000	86.7 (21.8)	94.2 (14.6)	97.2 (10.4)	96.5 (10.3)	91.7 (18.7)	6.4	4	p <.01
Dermatological problems								
1999	69.2 (26.2)	70.4 (26.5)	79.5 (24.1)	83.0 (22.8)	80.0 (27.1)	6.2	4	p <.001
2000	75.6 (27.9)	73.0 (26.4)	80.3 (21.9)	84.2 (20.7)	71.2 (28.5)	4.7	4	p <.01
2001	77.9 (24.8)	79.7 (25.6)	82.5 (23.1)	87.7 (18.1)	84.8 (19.5)	1.9	4	p <.01
Reproductive problems								
2000	85.9 (24.1)	87.4 (21.2)	90.0 (19.6)	92.1 (16.2)	82.8 (24.2)	2.8	4	p <.05
Psychological problems								
2000	83.7 (23.2)	84.6 (22.7)	87.2 (19.8)	89.5 (18.2)	80.2 (23.5)	2.6	4	p <.05
1999	76.6 (24.6)	78.7 (23.5)	82.1 (22.0)	86.9 (22.0)	85.7 (20.3)	3.0	4	p <.05
2001	78.6 (23.3)	86.1 (22.3)	85.9 (21.4)	94.6 (13.6)	88.4 (18.9)	4.8	4	p <.01

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients ||161≤ patients

Appendix B 20: Workload by genitourinary problems

	Scale scores					Statistical test		
	Mean* (Standard deviation)					ANOVA F	df	Sig.
	1†	2‡	3§	4¶	5			
Genitourinary problems								
1999	86.3 (21.0)	91.6 (18.1)	93.5 (14.4)	95.7 (12.2)	96.1 (10.9)	4.3	4	p <.01
2001	88.1 (20.5)	95.2 (14.2)	94.7 (12.8)	97.1 (11.1)	95.7 (11.4)	3.6	4	p <.01

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients ||161≤ patients

Appendix B 21: Workload by neurological problems to emergency medicine

	Scale scores Mean* (Standard deviation)					Statistical test		
	1†	2‡	3§	4¶	5	ANOVA F	df	Sig.
Neurological problems								
2000	79.3 (22.8)	80.3 (22.5)	82.9 (19.0)	88.2 (19.4)	78.1 (23.6)	2.9	4	$p<.05$
2001	78.6 (21.5)	82.4 (21.0)	82.5 (21.5)	89.7 (16.5)	86.2 (19.3)	2.8	4	$p<.05$
Sexually transmitted diseases								
1999	84.8 (23.4)	86.3 (23.5)	91.3 (16.6)	91.5 (18.3)	92.2 (16.5)	2.9	4	$p<.05$
2000	84.4 (23.1)	88.9 (20.5)	93.4 (16.1)	90.4 (17.1)	83.7 (22.7)	5.1	4	$p<.01$
2001	88.7 (21.3)	93.9 (16.4)	90.1 (19.7)	97.1 (11.1)	90.6 (15.2)	2.8	4	$p<.05$
Major trauma and disasters								
1999	65.2 (27.3)	68.2 (26.6)	72.7 (27.8)	72.3 (27.1)	81.9 (26.0)	2.9	4	$p<.05$
2000	63.0 (25.8)	63.1 (28.6)	68.0 (26.8)	78.1 (27.0)	68.4 (25.8)	4.1	4	$p<.01$
Emergency medicine								
1999	81.9 (24.1)	81.8 (22.7)	87.4 (20.2)	85.8 (21.6)	95.1 (12.0)	3.8	4	$p<.01$
2000	78.5 (25.8)	81.3 (24.5)	87.2 (21.5)	89.5 (19.0)	86.1 (20.9)	3.2	4	$p<.05$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients ||161≤ patients

Appendix B 22: Workload by sexually transmitted diseases to public health

	Scale scores Mean* (Standard deviation)					Statistical test		
	1†	2‡	3§	4¶	5	ANOVA F	df	Sig.
Sexually transmitted diseases								
1999	84.8 (23.4)	86.3 (23.5)	91.3 (16.6)	91.5 (18.3)	92.2 (16.5)	2.9	4	p < .05
2000	84.4 (23.1)	88.9 (20.5)	93.4 (16.1)	90.4 (17.1)	83.7 (22.7)	5.1	4	p < .01
2001	88.7 (21.3)	93.9 (16.4)	90.1 (19.7)	97.1 (11.1)	90.6 (15.2)	2.8	4	p < .05
Major trauma and disasters								
1999	65.2 (27.3)	68.2 (26.6)	72.7 (27.8)	72.3 (27.1)	81.9 (26.0)	2.9	4	p < .05
2000	63.0 (25.8)	63.1 (28.6)	68.0 (26.8)	78.1 (27.0)	68.4 (25.8)	4.1	4	p < .01
Emergency medicine								
1999	81.9 (24.1)	81.8 (22.7)	87.4 (20.2)	85.8 (21.6)	95.1 (12.0)	3.8	4	p < .01
2000	78.5 (25.8)	81.3 (24.5)	87.2 (21.5)	89.5 (19.0)	86.1 (20.9)	3.2	4	p < .05
Preventive medicine								
1999	90.1 (18.4)	92.5 (16.4)	96.2 (11.1)	96.8 (11.0)	96.2 (10.8)	4.2	4	p < .01
2000	92.6 (15.7)	95.7 (12.6)	97.4 (9.6)	96.1 (13.3)	91.0 (18.4)	4.3	4	p < .01
Public health medicine								
2000	81.5 (24.2)	87.4 (19.5)	89.5 (17.6)	89.5 (18.2)	79.5 (22.9)	5.3	4	p < .01

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients ||161≤ patients

Appendix B 23: Locality by particular patient presentations

	Scale scores		Statistical test		
	Mean* (Standard deviation)		<i>t</i> test	<i>df</i>	Sig.
	Urban areas	Rural and remote areas			
Major trauma and disasters					
1999	69.0 (27.2)	75.0 (28.4)	-2.4	344	<i>p</i> <.05
2000	65.3 (26.6)	71.6 (28.0)	-2.2	283	<i>p</i> <.05
2001	68.4 (25.9)	79.5 (25.4)	-4.3	443	<i>p</i> <.001
Emergency medicine					
1999	83.9 (22.5)	87.7 (20.1)	-2.0	397	<i>p</i> <.05
2001	85.7 (20.9)	90.6 (19.0)	-2.5	326	<i>p</i> <.05
Sexually transmitted diseases					
2000	85.9 (21.6)	90.2 (19.3)	-2.1	323	<i>p</i> <.05
Drug and alcohol problems					
2000	73.4 (23.9)	78.6 (23.9)	-2.0	392	<i>p</i> <.05

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 24: Longitudinal findings for particular patient presentations

	Scale scores Mean* (Standard deviation)			Statistical test		
	1999	2000	2001	ANOVA F	df	Sig.
Eye problems	71.2 (26.4)	85.5 (19.4)	90.2 (17.4)	28.7	2	$p<.001$
Ear, nose and throat problems	82.6 (23.1)	93.3 (15.2)	96.7 (10.9)	23.7	2	$p<.001$
Respiratory problems	98.0 (8.0)	100.0 (0.0)	99.8 (2.9)	4.1	2	$p<.05$
Cardiovascular problems	92.6 (15.1)	98.2 (7.6)	98.5 (7.0)	9.9	2	$p<.001$
Dermatological problems	64.6 (24.8)	76.9 (22.7)	88.2 (17.5)	56.8	2	$p<.001$
Gastrointestinal problems	95.4 (12.2)	97.5 (9.8)	99.5 (4.1)	7.3	2	$p<.01$
Musculoskeletal problems	77.3 (24.6)	84.0 (22.9)	91.2 (16.9)	30.3	2	$p<.001$
Genitourinary problems	89.1 (19.6)	95.1 (12.5)	98.2 (7.5)	15.1	2	$p<.001$
Reproductive problems	81.4 (24.2)	87.8 (20.8)	92.4 (16.3)	11.6	2	$p<.001$
Neurological problems	77.1 (22.7)	84.6 (19.5)	87.7 (17.8)	12.5	2	$p<.001$
Sexually transmitted diseases	79.6 (24.0)	89.8 (18.4)	94.7 (13.6)	23.2	2	$p<.001$
Drug and alcohol problems	72.4 (25.0)	76.5 (23.3)	80.9 (23.4)	7.2	2	$p<.01$
Pediatrics	92.6 (16.2)	97.4 (9.8)	98.7 (7.7)	8.7	2	$p<.001$
HIV/AIDS problems	57.9 (25.8)	61.8 (23.9)	67.4 (24.4)	6.9	2	$p<.01$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 25: Longitudinal findings for preventive to public health medicine

	Scale scores			Statistical test		
	Mean* (Standard deviation)					
	1999	2000	2001	ANOVA F	df	Sig.
Psychological problems	78.5 (23.4)	86.7 (20.2)	90.5 (17.7)	17.8	2	$p<.001$
Preventive medicine	91.8 (17.1)	95.9 (11.8)	98.2 (7.6)	9.1	2	$p<.001$
Public health medicine	83.0 (20.8)	88.1 (19.4)	93.1 (14.7)	12.8	2	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

Appendix B 26: Stage of training by satisfaction with learning items

	Scale scores Mean* (Standard deviation)					Statistical test		
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
Range of clinical experiences								
1999	83.9 (21.4)	87.9 (19.0)	93.8 (14.8)	92.5 (16.9)	91.4 (17.6)	5.5	4	<i>p</i> <.001
2000	86.1 (20.2)	94.0 (16.5)	92.8 (16.1)	95.3 (13.0)	92.4 (17.3)	4.3	4	<i>p</i> <.01
2001	88.1 (19.7)	90.1 (20.6)	92.7 (15.3)	95.9 (12.5)	95.2 (13.6)	3.7	4	<i>p</i> <.01
Learning challenges provided								
2000	88.5 (15.9)	91.3 (17.4)	96.4 (10.4)	95.6 (12.7)	88.8 (20.4)	5.2	4	<i>p</i> <.001
Role models provided by supervisors								
1999	74.6 (22.3)	89.8 (20.4)	88.2 (19.6)	88.0 (19.1)	79.8 (26.7)	8.9	4	<i>p</i> <.001
2000	79.2 (18.8)	91.0 (19.3)	90.3 (19.3)	93.8 (16.9)	82.9 (23.4)	9.8	4	<i>p</i> <.001
2001	70.2 (20.2)	91.9 (16.5)	87.9 (20.7)	91.0 (17.1)	87.5 (20.6)	10.6	4	<i>p</i> <.001

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

[†]Hospital trainees [‡]Basic trainees [§]Advanced trainees [¶]Subsequent trainees ^{||}Former trainees

Appendix B 27: Stage of training by satisfaction with supervision items

	Scale scores Mean* (Standard deviation)					Statistical test		
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
Supervision and advice received								
1999	73.4 (22.3)	90.6 (19.1)	91.1 (18.6)	85.8 (21.7)	79.4 (26.4)	10.2	4	<i>p</i> <.001
2000	80.7 (19.6)	89.7 (19.7)	84.4 (23.0)	94.0 (15.3)	80.5 (25.1)	9.7	4	<i>p</i> <.001
2001	75.3 (22.1)	94.6 (12.5)	87.3 (21.7)	89.7 (18.5)	86.4 (21.6)	6.1	4	<i>p</i> <.001
Opportunities to learn communication skills								
2000	84.4 (20.5)	88.1 (20.7)	94.4 (13.9)	93.5 (15.6)	88.4 (19.7)	4.5	4	<i>p</i> <.01
2001	83.7 (21.5)	85.6 (21.0)	89.1 (18.2)	93.6 (13.8)	91.8 (17.6)	4.1	4	<i>p</i> <.01
Access to educational materials and resources								
1999	79.1 (22.5)	87.1 (20.6)	86.7 (21.7)	88.5 (18.0)	86.7 (22.1)	3.4	4	<i>p</i> <.05
2001	79.1 (21.0)	79.3 (25.3)	84.2 (23.9)	91.0 (16.0)	89.6 (19.7)	5.7	4	<i>p</i> <.001

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ¶Subsequent trainees ||Former trainees

Appendix B 28: Stage of training by further satisfaction items

	Scale scores Mean* (Standard deviation)					Statistical test		
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
Opportunities to learn about population health in the context of general practice								
1999	66.7 (21.0)	76.0 (22.8)	78.0 (23.4)	78.7 (20.9)	73.3 (23.6)	4.8	4	<i>p</i> <.01
2000	69.6 (23.4)	75.0 (22.6)	81.0 (25.0)	82.2 (21.1)	75.6 (25.8)	4.4	4	<i>p</i> <.01
2001	67.3 (23.1)	74.8 (24.1)	75.2 (22.4)	82.1 (21.2)	80.3 (21.8)	4.9	4	<i>p</i> <.01
Opportunities to meet with other trainees								
1999	74.2 (24.2)	86.6 (22.1)	90.2 (18.8)	85.9 (22.1)	87.7 (19.9)	9.1	4	<i>p</i> <.001
2000	73.5 (24.8)	89.8 (19.7)	88.2 (21.6)	89.1 (19.5)	92.4 (16.9)	13.0	4	<i>p</i> <.001
2001	72.4 (27.8)	81.1 (92.4)	83.6 (23.0)	86.4 (21.4)	91.1 (19.2)	8.4	4	<i>p</i> <.001
External clinical teaching sessions								
1999	71.7 (24.0)	84.4 (22.3)	88.3 (23.0)	84.8 (22.6)	83.0 (23.8)	7.3	4	<i>p</i> <.001
2001	72.0 (21.7)	94.3 (17.1)	86.7 (22.8)	85.5 (22.0)	85.4 (23.0)	5.9	4	<i>p</i> <.001

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ¶Subsequent trainees ||Former trainees

Appendix B 29: Stage of training by miscellaneous satisfaction items

	Scale scores Mean* (Standard deviation)					Statistical test		
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
Video and audio consultation reviews								
2001	63.2 (15.7)	72.9 (24.6)	66.0 (24.2)	72.4 (23.4)	71.5 (24.2)	2.9	4	<i>p</i> <.05
Usefulness of the trainee logbook								
1999	66.0 (20.6)	57.0 (21.2)	55.7 (21.1)	50.8 (20.2)	52.2 (22.7)	8.6	4	<i>p</i> <.01
2000	67.5 (21.2)	61.5 (22.3)	54.9 (23.1)	57.5 (24.7)	55.1 (23.3)	4.7	4	<i>p</i> <.001
Location of training posts								
1999	70.8 (23.7)	85.7 (23.0)	84.3 (25.0)	84.0 (24.8)	83.7 (24.8)	6.8	4	<i>p</i> <.001
2001	75.8 (22.2)	82.0 (26.8)	83.6 (24.7)	89.7 (19.4)	84.6 (23.7)	3.7	4	<i>p</i> <.01
Remuneration and allowances received								
1999	63.2 (23.3)	68.1 (25.6)	73.6 (27.3)	80.5 (25.1)	69.9 (27.1)	6.9	4	<i>p</i> <.001
2000	67.6 (23.4)	70.9 (28.3)	64.1 (27.8)	78.7 (26.3)	67.1 (27.0)	4.9	4	<i>p</i> <.01
2001	65.3 (26.0)	72.2 (28.2)	69.7 (27.4)	80.8 (25.4)	72.0 (28.7)	3.8	4	<i>p</i> <.01

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ¶Subsequent trainees ||Former trainees

Appendix B 30: Stage of training by satisfaction with administrative items

	Scale scores Mean* (Standard deviation)					Statistical test		
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
Other terms and conditions of employment								
1999	66.7 (19.7)	74.9 (25.4)	78.5 (25.6)	83.2 (24.3)	74.1 (25.9)	6.7	4	<i>p</i> <.001
2000	71.4 (22.1)	81.7 (24.5)	76.9 (24.9)	84.9 (21.0)	71.1 (26.5)	7.9	4	<i>p</i> <.001
2001	65.3 (23.3)	79.3 (24.0)	77.6 (26.5)	85.9 (21.5)	79.7 (25.6)	6.6	4	<i>p</i> <.001
Information sent to you by the program								
1999	84.4 (21.6)	80.4 (23.0)	89.1 (19.2)	86.6 (21.2)	83.2 (21.7)	2.5	4	<i>p</i> <.05
2000	83.1 (22.9)	91.7 (16.6)	88.2 (18.1)	90.5 (18.1)	80.9 (24.2)	6.4	4	<i>p</i> <.001
2001	71.2 (28.8)	83.8 (21.7)	86.1 (20.0)	88.2 (18.5)	86.2 (21.8)	6.2	4	<i>p</i> <.001
Administration of training								
1999	79.2 (24.8)	76.0 (25.7)	82.8 (24.5)	82.3 (24.9)	73.2 (27.2)	2.9	4	<i>p</i> <.05
2000	80.2 (22.6)	85.1 (20.7)	78.5 (23.9)	85.4 (21.4)	75.6 (27.0)	4.2	4	<i>p</i> <.01
2001	71.9 (27.8)	75.7 (26.8)	84.3 (25.0)	89.2 (18.2)	83.4 (23.8)	6.1	4	<i>p</i> <.001

**Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ¶Subsequent trainees ||Former trainees

Appendix B 31: Stage of training by satisfaction with level of workload

	Scale scores Mean* (Standard deviation)					Statistical test		
	H [†]	B [‡]	A [§]	S	F [¶]	ANOVA F	df	Sig.
Level of workload expected								
1999	81.4 (17.0)	87.4 (20.5)	91.2 (16.4)	91.4 (16.9)	89.0 (19.6)	5.1	4	<i>p</i> <.001
2000	80.7 (20.2)	89.2 (19.9)	83.1 (24.4)	93.1 (16.3)	87.1 (20.7)	6.0	4	<i>p</i> <.001
2001	76.9 (20.6)	91.9 (16.5)	88.9 (19.4)	94.6 (12.3)	92.5 (16.9)	11.2	4	<i>p</i> <.001

**Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†Hospital trainees ‡Basic trainees §Advanced trainees ¶Subsequent trainees ||Former trainees

Appendix B 32: Gender by satisfaction

	Scale scores		Statistical test		
	Mean* (Standard deviation)				
	Male	Female	<i>t</i> test	<i>df</i>	Sig.
Opportunities to learn about population health					
2001	80.9 (21.3)	76.6 (23.1)	2.2	463	<i>p</i> <.05
Access to educational materials and resources					
2000	85.9 (20.6)	89.6 (19.0)	-2.1	441	<i>p</i> <.05
Location of training posts					
1999	80.0 (25.2)	84.1 (23.7)	-2.0	584	<i>p</i> <.05
Remuneration and allowances received					
1999	67.7 (27.7)	73.9 (25.7)	-2.8	583	<i>p</i> <.01
Other terms and conditions of employment					
1999	73.1 (25.9)	77.5 (24.1)	-2.1	580	<i>p</i> <.05
2000	74.4 (25.6)	79.7 (23.8)	-2.5	525	<i>p</i> <.05
Video and consultation reviews					
2000	72.8 (24.4)	67.8 (23.4)	2.3	433	<i>p</i> <.05

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are satisfied with training: 1=strongly disagree, 100=strongly agree.

Appendix B 33: Locality by satisfaction

	Scale scores Mean* (Standard deviation)		Statistical test		
	Urban areas	Rural and remote areas	<i>t</i> test	df	Sig.
Access to educational materials and resources					
1999	87.1 (19.7)	82.7 (23.7)	2.2	303	<i>p</i> <.05
2000	90.8 (16.6)	83.9 (23.0)	3.2	230	<i>p</i> <.01
2001	89.1 (18.5)	84.3 (22.7)	2.2	254	<i>p</i> <.05
Range of clinical experiences					
2000	91.2 (17.5)	94.9 (14.4)	-2.2	344	<i>p</i> <.05
Location of training posts					
2000	80.5 (24.2)	86.9 (21.7)	-2.7	325	<i>p</i> <.01
2001	82.4 (24.2)	89.6 (19.3)	-3.4	367	<i>p</i> <.01

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are satisfied with training: 1=strongly disagree, 100=strongly agree.

Appendix B 34: Workload by satisfaction

	Scale scores					Statistical test		
	Mean (Standard deviation)					ANOVA F	df	Sig.
	1†	2‡	3§	4¶	5			
Range of clinical experiences								
1999	83.8 (22.7)	89.4 (17.4)	92.8 (16.1)	91.2 (19.8)	92.2 (14.4)	3.5	4	<i>p</i> <.01
2000	87.5 (23.2)	93.1 (15.9)	93.8 (16.2)	95.2 (13.0)	88.8 (17.9)	3.0	4	<i>p</i> <.05
2001	87.3 (20.8)	93.2 (16.2)	95.0 (13.1)	95.2 (14.2)	98.6 (6.9)	4.3	4	<i>p</i> <.01
Role models provided by supervisors								
2000	84.2 (20.0)	90.2 (18.8)	90.1 (19.9)	87.3 (22.4)	81.5 (21.2)	3.6	4	<i>p</i> <.01
Opportunities to learn about population health								
1999	66.2 (23.7)	75.4 (24.1)	75.8 (21.5)	77.6 (21.7)	77.4 (26.9)	3.0	4	<i>p</i> <.05
2001	82.7 (23.8)	90.8 (18.1)	93.1 (15.0)	91.9 (15.5)	92.0 (14.4)	3.9	4	<i>p</i> <.01
Opportunities to meet with other trainees								
2000	86.5 (22.2)	88.5 (20.6)	90.4 (19.5)	88.6 (19.3)	81.6 (23.2)	3.0	4	<i>p</i> <.05

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients ||161≤ patients

Appendix B 35: Workload by further satisfaction items

	Scale scores Mean (Standard deviation)					Statistical test		
	1†	2‡	3§	4¶	5	ANOVA F	df	Sig.
Opportunities to learn communication skills								
2001	82.7 (23.8)	90.8 (18.1)	93.1 (15.0)	91.9 (15.5)	92.0 (14.4)	3.9	4	$p<.01$
Video and audio consultation reviews								
2001	64.8 (24.6)	74.7 (23.6)	69.5 (24.5)	76.8 (23.8)	77.8 (23.6)	3.3	4	$p<.05$
External clinical teaching sessions								
2000	85.0 (21.3)	86.3 (22.3)	87.1 (21.7)	86.0 (20.6)	74.8 (24.1)	5.3	4	$p<.01$
Location of training posts								
2000	85.4 (22.4)	80.3 (24.5)	85.9 (22.4)	87.3 (21.1)	77.5 (26.2)	3.1	4	$p<.05$
Other terms and conditions of employment								
2000	82.1 (21.1)	81.8 (22.5)	78.6 (24.3)	75.4 (28.0)	70.0 (24.4)	3.8	4	$p<.01$
2001	75.0 (27.9)	85.0 (22.4)	84.3 (24.0)	87.9 (21.4)	85.5 (20.7)	2.7	4	$p<.05$
Information sent out to trainees								
2001	76.2 (26.0)	85.4 (20.9)	87.2 (20.5)	90.5 (19.8)	87.7 (19.1)	4.0	4	$p<.01$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are competent to practise unsupervised general practice: 1=strongly disagree, 100=strongly agree.

†1–40 patients ‡41–80 patients §81–120 patients ¶121–160 patients ||161+ patients

Appendix B 36: Longitudinal findings for satisfaction items

	Scale scores			Statistical test		
	Mean* (Standard deviation)			ANOVA F	df	Sig.
Range of clinical experiences	86.9 (19.2)	94.7 (14.2)	96.2 (11.4)	13.9	2	$p<.001$
Learning challenges provided	90.1 (17.4)	94.7 (13.6)	96.7 (10.8)	6.5	2	$p<.01$
Role models provided by supervisors	84.3 (22.2)	92.3 (17.8)	91.6 (17.0)	8.1	2	$p<.001$
Supervision and advice received	85.1 (21.0)	91.9 (17.3)	90.6 (17.4)	4.1	2	$p<.05$
Opportunities to learn about communication skills	86.8 (21.1)	93.7 (16.7)	93.7 (14.4)	7.0	2	$p<.01$
Access to educational materials and resources	84.8 (21.7)	89.8 (18.1)	92.7 (15.1)	8.0	2	$p<.001$
Opportunities to learn about population health in the context of general practice	73.7 (23.5)	78.9 (23.1)	85.9 (20.5)	10.3	2	$p<.001$
Opportunities to meet with other trainees	82.6 (22.8)	90.0 (19.8)	90.3 (18.8)	6.5	2	$p<.01$
External clinical teaching sessions	81.6 (22.3)	87.5 (20.2)	88.3 (20.9)	4.9	2	$p<.01$
Video and consultation reviews	67.0 (21.3)	71.1 (23.5)	78.6 (23.1)	11.0	2	$p<.01$
Location of training posts	78.7 (24.5)	86.0 (21.7)	85.7 (22.6)	5.8	2	$p<.01$
Remuneration and allowances received	67.5 (27.1)	75.3 (27.5)	77.8 (26.0)	7.4	2	$p<.01$
Other terms and conditions of employment	74.4 (24.3)	82.4 (23.2)	86.0 (22.7)	9.5	2	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are satisfied with training: 1=strongly disagree, 100=strongly agree.

Appendix B 37: Longitudinal findings for satisfaction with information items

	Scale scores			Statistical test		
	Mean* (Standard deviation)			ANOVA F	df	Sig.
1999	2000	2001				
Information sent to you by the program	84.2 (21.3)	91.9 (16.5)	90.2 (16.9)	9.8	2	$p<.001$
Administration of training	79.6 (25.1)	85.5 (21.2)	89.5 (18.2)	7.5	2	$p<.01$
Level of workload expected	78.1 (12.7)	92.5 (16.2)	93.8 (15.4)	17.9	2	$p<.001$

*Mean on a one hundred-point scale represents the extent to which respondents agree that they are satisfied with training: 1=strongly disagree, 100=strongly agree.

Appendix C

Appendix C 1: Questionnaire items as they relate to five domains of general practice

To what extent do you agree or disagree that you are competent in the following unsupervised general practice activities?

Domain one: communication skills and the patient-doctor relationship

- Provide continuity of care using a whole-person approach
- Communicate with patients effectively
- Counsel patients effectively

Domain two: applied professional knowledge and skills

- Manage a wide range of health problems and medical conditions
- Apply principles of primary health care
- Critically evaluate relevant information
- Order cost-effective investigations
- Make concise, clear and appropriate referrals

Domain three: population health and the context of general practice

- Offer integrated care
- Undertake a public health role
- Function effectively in the health care system
- Understand the relationship between patterns and prevalence of disease and general practice
- Make clinical decisions based on the best available evidence

Domain four: professional and ethical role

- Provide support and advice on health and illness in an ethical manner
- Establish your own professional support networks
- Engage in on-going, self-directed learning and professional development

Domain five: organizational and legal dimensions

- Apply relevant legal principles to own general practice
- Implement effective practice management systems
- Effectively maintain and use patient records

Appendix D

Appendix D 1: General practice vocational training questionnaire



The Royal Australian College of General Practitioners

Outcomes Evaluation Unit (OEU)

Registrar Views on Outcomes of Training

C U R R E N T A N D F O R M E R R E G I S T R A R Q U E S T I O N N A I R E

Please complete all questions in the questionnaire.

*Please place the completed questionnaire in the reply paid envelope provided
and return to the OEU.*

Section One

Areas of competence

Please **circle** the most appropriate response to each question.

Strongly disagree		Neutral		Strongly agree
	Disagree		Agree	
1	2	3	4	5

1 To what extent do you agree or disagree that you are competent in the following unsupervised general practice activities?

- Provide continuity of care using a whole-person approach
- Manage a wide range of health problems and medical conditions
- Communicate with patients effectively
- Counsel patients effectively
- Apply principles of primary health care
- Critically evaluate relevant information
- Order cost-effective investigations
- Make clear, concise and appropriate referrals
- Offer integrated care
- Undertake a public health role as a GP
- Function effectively in the health care system
- Provide support and advice on health and illness in an ethical manner
- Establish your own professional support networks
- Engage in ongoing, self-directed learning and professional development
- Implement effective practice management systems
- Apply relevant legal principles to your general practice
- Effectively maintain and use patient records
- Understand the relationship between patterns and prevalence of disease and general practice
- Make clinical decisions based on the best available evidence

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

Please **circle** the most appropriate response to each question.

Strongly disagree		Neutral		Strongly agree
	Disagree		Agree	
1	2	3	4	5

2 To what extent do you agree or disagree that you are competent in the unsupervised management of the following presenting problems?

a. Eye problems	1	2	3	4	5
b. Ear, nose and throat problems	1	2	3	4	5
c. Respiratory problems	1	2	3	4	5
d. Cardiovascular problems	1	2	3	4	5
e. Gastrointestinal and digestive problems	1	2	3	4	5
f. Dermatological problems	1	2	3	4	5
g. Musculoskeletal problems	1	2	3	4	5
h. Genito-urinary problems	1	2	3	4	5
i. Reproductive problems	1	2	3	4	5
j. Neurological problems	1	2	3	4	5
k. Sexually-transmitted diseases	1	2	3	4	5
l. Drug and alcohol problems	1	2	3	4	5
m. Paediatrics	1	2	3	4	5
n. HIV/AIDS problems	1	2	3	4	5
o. Major trauma and disasters	1	2	3	4	5
p. Emergency medicine	1	2	3	4	5
q. Psychological problems	1	2	3	4	5
r. Preventive medicine	1	2	3	4	5
s. Public health medicine	1	2	3	4	5

Please **circle** the most appropriate response to each question.

Strongly disagree		Neutral		Strongly agree
	Disagree		Agree	
1	2	3	4	5

3 To what extent do you agree or disagree that you are competent in the unsupervised medical care of the following groups of patients?

a. Neonates	1	2	3	4	5
b. Infants	1	2	3	4	5
c. Children	1	2	3	4	5
d. Adolescents	1	2	3	4	5
e. Women	1	2	3	4	5
f. Men	1	2	3	4	5
g. Older people	1	2	3	4	5
h. Indigenous people	1	2	3	4	5
i. People from non-English speaking backgrounds	1	2	3	4	5
j. People from remote rural communities	1	2	3	4	5
k. People with mental illness	1	2	3	4	5
l. People with disabilities	1	2	3	4	5
m. People with chronic conditions	1	2	3	4	5

Section Two

Satisfaction with aspects of your training

Please **circle** the most appropriate response to each question.

Very dissatisfied		Neutral		Very satisfied
1	2	3	4	5

1 How satisfied are you with the following?

a. Your range of clinical experiences	1	2	3	4	5
b. The learning challenges provided	1	2	3	4	5
c. The role model provided by your supervisor for general practice	1	2	3	4	5
d. The supervision and advice received	1	2	3	4	5
e. Opportunities to learn communication skills	1	2	3	4	5
f. Access to educational materials and resources	1	2	3	4	5
g. Opportunities to learn about population health in the context of general practice	1	2	3	4	5
h. Opportunities to meet with other registrars	1	2	3	4	5
i. External clinical teaching sessions	1	2	3	4	5
j. Video and audio consultation reviews	1	2	3	4	5
k. Usefulness of the registrar logbook	1	2	3	4	5
l. Location of your training placements	1	2	3	4	5
m. The remuneration and allowances received	1	2	3	4	5
n. Other terms and conditions of employment	1	2	3	4	5
o. Information sent to you about your training	1	2	3	4	5
p. Administration of training	1	2	3	4	5
q. The level of workload expected	1	2	3	4	5
r. Opportunities to gain skills in practice management	1	2	3	4	5

Section Three

Participation in training

*If you are currently a **basic GP term registrar**, an **advanced GP term registrar**, or a registrar doing **subsequent GP experience**, continue with question 1 below.*

*If you are currently a **hospital registrar in the hospital year** or a **former registrar**, please go to section 4, question 1, on page 7.*

*The following questions ask about the **last term** you completed in your training program. Please **circle** the most appropriate response for each.*

1 Which term was the *last* one you completed?

- 1 Hospital term
(go to section 4, question 1)
- 2 Basic GP term
- 3 Advanced GP term
- 4 Subsequent GP experience

2 When did you complete this term?

- 1 Within the last month
- 2 Within the last two months
- 3 Within the last three months
- 4 More than three months ago

3 How many hours on average did you spend in clinical practice each week during your last term?

- 1 hours
- 2 Not applicable

Please **circle** the most appropriate response for each of the following questions about the most recent term you completed.

4 What percentage of the time were you and a supervisor in the practice at the same time?

- 1 Not at all
- 2 per cent
- 3 Not applicable

5 What percentage of the time was a supervisor available by phone while you were in the practice?

- 1 Not at all
- 2 per cent
- 3 Not applicable

6 What percentage of the time was a supervisor available by phone while you were on call?

- 1 Not at all
- 2 per cent
- 3 Not applicable

7 Educational releases are workshops, half-day sessions or educational teleconferences conducted during training. Approximately how many hours in total of educational release did you attend during your last term?

- 1 None
- 2 hours
- 3 Not applicable

8 How many times in a typical fortnight did you receive educational advice or teaching from a supervisor?

- 1 None
- 2 times
- 3 Not applicable

9 How many times over a typical fortnight did your supervisor observe your clinical consultations?

- 1 None
- 2 times
- 3 Not applicable

Section Four

Demographic information

Please complete your response for each of the following questions.

1 What was the year of your first enrolment in the Training Program?

— — — —

2 What term of training are you currently in?

- 1 Hospital
- 2 Basic GP
- 3 Advanced GP
- 4 Subsequent GP experience
- 5 Former – Completed Training Program

3 What is your gender?

- 1 Male
- 2 Female

4 During your last term of your training, were you:

- 1 Part-time
- 2 Full-time
- 3 On leave

5 How old are you?

— — years

6 Where is your current main practice located?

- 1 Metropolitan area
- 2 Rural area
- 3 Remote area
- 4 Other – *please specify*

7 Is this practice in an Area of Medical Service Need?

- 1 Yes
- 2 No
- 3 Not applicable
- 4 Don't know

8 In which state or territory do you currently practise medicine?

- 1 NSW
- 2 Vic.
- 3 Qld
- 4 WA
- 5 SA/NT
- 6 ACT
- 7 Nth Qld
- 8 Tas.
- 9 Outside Australia
- 10 Not applicable

9 In which of the following countries were you born?

- 1 Australia
- 2 England
- 3 Scotland
- 4 Italy
- 5 Greece
- 6 New Zealand
- 7 Viet Nam
- 8 Other – *please specify*

10 What is the primary language used at home?

- 1 English
- 2 Italian
- 3 Greek
- 4 Vietnamese
- 5 Other – *please specify*

11 Which components of training have you completed?
(Please circle each applicable answer.)

- 1 Hospital terms
- 2 Basic GP terms
- 3 Advanced terms
- 4 Subsequent GP experience
- 5 Special skills training
- 6 Advanced rural skills posts
- 7 Advanced academic posts
- 8 Other – please specify

12 Are/were you in the rural training stream?

- 1 Yes
- 2 No
- 3 Not applicable

13 What is your main occupation?
(Circle one answer only.)

- 1 GP registrar
- 2 Solo GP
- 3 General practice principal/partner
- 4 Group practice employee
- 5 Locum GP
- 6 Community health centre GP
- 7 Rural hospital medical officer
- 8 Medical educator
- 9 Academic GP
- 10 Clinical researcher
- 11 Other – please specify

14 What is the average number of patients you see over a week?

- 1 None
- 2 patients
- 3 Not applicable

15 At which university did you complete your undergraduate medical degree?

Please specify

16

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Thank you for your cooperation.

Please place the completed questionnaire in the reply paid envelope provided and return to:

REPLY PAID 69633
Sue Goldman
RACGP
1 Palmerston Crescent
SOUTH MELBOURNE VIC 3205

If you have any questions about this questionnaire please contact:

Ms Sue Goldman
Evaluation Officer
Outcomes Evaluation Unit
Telephone: (03) 9214 1554
Fax: (03) 9214 1570
e-mail: sue.goldman@racgp.org.au

